

STIC-Biotech/ChemLib

192484

From: Li, Ruixiang  
Sent: Friday, June 09, 2006 2:23 PM  
To: STIC-Biotech/ChemLib  
Subject: Sequence search of Application No.10/644,084

Please do a standard search on:

- (i). SEQ ID NO: 2 against amino acid databases (excluding pending databases).
- (ii). SEQ ID NOS: 1 and 2 against nucleic acid databases (excluding pending databases).

Thank you very much!

Ruixiang Li  
GAU 1646  
REM 4D75  
Mail Box 4C70  
(571) 272-0875

78733

URFE

\*\*\*\*\*

Searcher: \_\_\_\_\_  
Searcher Phone: \_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date completed: \_\_\_\_\_  
Searcher Prep Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA# \_\_\_\_\_ AA# \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure #: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (Specify): \_\_\_\_\_

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 20:46:55 ; Search time 11414 Seconds  
(without alignments)  
13188.614 Million cell updates/sec

Title: US-10-644-084-1

Perfect score: 2692

Sequence: 1 cgtagagagtacagagc.....ttgattatggagtaatgggg 2692

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780: residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est3:\*

3: gb\_est4:\*

4: gb\_est5:\*

5: gb\_est6:\*

6: gb\_hc:\*

7: gb\_est2:\*

8: gb\_est7:\*

9: gb\_est8:\*

10: gb\_est9:\*

11: gb\_gse1:\*

12: gb\_gse2:\*

13: gb\_gse3:\*

14: gb\_gse4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2676	99.4	3185	AK043865	AK043865 Mus muscu
2	2676	99.4	3413	AK031356	AK031356 Mus muscu
3	1985.4	73.8	2245	AK077112	AK077112 Mus muscu
4	1973	73.3	2371	AK049080	AK049080 Mus muscu
5	1832	68.1	1845	AY412493	AY412493 Mus muscu
6	1701.4	63.2	3154	BC035580	BC035580 Homo sapi
7	1701.4	63.2	5658	HS0801317	HS0801317 Homo sapi
8	1416.6	52.6	2111	CR613292	CR613292 full-leng
9	1363.8	50.7	1845	AY412491	AY412491 Homo sapi
10	1357.4	50.4	1845	AY412492	AY412492 Pan trogl
11	876.4	32.6	878	AY412492	AY412492 Pan trogl
12	812	30.2	840	CF949831	CF949831 UI-M-IB0-
13	810	30.1	834	CF949831	CF949831 UI-M-IB0-
14	802	29.8	804	CM535911	CM535911 UI-M-IB0-
15	798	29.6	798	CM535911	CM535911 UI-M-IB0-
16	789	29.3	822	CM562859	CM562859 UI-M-IB0-
17	768.4	28.5	771	CM563162	CM563162 UI-M-IB0-
18	761	28.3	933	BQ179212	BQ179212 UI-M-IB0-
19	753.2	28.0	771	CF737129	CF737129 UI-M-IB0-

20 750.2 27.9 770 5 CF729394  
21 749 27.8 760 8 CN528686  
22 748.2 27.8 845 3 BQ769157  
23 739.2 27.5 746 8 CN528831  
24 721.4 26.8 723 5 CF737957  
25 710.4 26.4 712 5 CF732776  
26 690 25.6 736 8 CX226608  
27 679 25.2 679 8 CN532570  
28 677.8 25.2 683 8 CO424815  
29 675 25.1 687 9 CX568194  
30 658 24.4 903 10 DV893190  
31 653.2 24.3 1005 1 AL582610  
32 643 23.9 643 2 BG085173  
33 638 23.7 640 5 CF737822  
34 633 23.5 833 8 CV117234  
35 623.4 23.2 638 8 CN720771  
36 622.6 23.1 954 1 AL561392  
37 622.2 23.1 635 5 CF899025  
38 616.4 22.9 860 10 DT831879  
39 614.4 22.8 633 1 AV279314  
40 614.2 22.8 805 2 BG242849  
41 614.2 22.8 828 10 DV803752  
42 612.2 22.7 658 7 BB614736  
43 606.8 22.5 623 4 CA548100  
44 601.2 22.3 752 7 BF539334  
45 600 22.3 600 2 BG804175

## ALIGNMENTS

RESULT 1  
AK043865

LOCUS  
DEFINITION

AK043865 3185 bp mRNA linear HTC 02-SEP-2005  
Mus musculus 10 days neonate cortex cDNA, RIKEN full-length  
enriched library, clone:A830043F14 product:HYPOTHETICAL 71.0 KDA  
PROTEIN homolog [Mus musculus], full insert sequence.

ACCESSION

VERSION AK043865.1 GI:26335971

KEYWORDS  
HTC; CAP trapper.

SOURCE  
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

**HIS PAGE LEFT BLANK**





THIS PAGE LEFT BLANK



THIS PAGE LEFT BLANK



**THIS PAGE LEFT BLANK**



THIS PAGE LEFT BLANK.



Db 2773 ATAAGTGAAGTACTAAGCTTACATTTTGTGAGATTTTAAAGCATTTGATTTATTTTAT 2832  
 Qy 2641 ATATGTGAATGTATAAATTTCTTAAGAGGAATATTGATTTATGAGTAATGAGTAAATGGGG 2692  
 Db 2833 ATATGTGAATGTATAAATTTCTTAAGAGGAATATTGATTTATGAGTAATGAGTAAATGGGG 2884

RESULT 3  
 BC021749 3425 bp mRNA linear ROD 18-JUL-2005  
 LOCUS Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,  
 DEFINITION mRNA (CDNA clone MGC:25823 IMAGE:4165430), complete cds.  
 ACCESSION BC021749  
 VERSION BC021749.1 GI:18256805  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Iqbal, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Mammalian Gene Collection Program Team  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16999-16903 (2002)  
 12477932  
 2 (bases 1 to 3425)

REFERENCE  
 AUTHORS NIH MGC Project  
 Direct Submission  
 Submitted (18-JAN-2002) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

REFERENCE  
 AUTHORS Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegue, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

FEATURES  
 source  
 1..3425  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 30 Row: m Column: 16.  
 Location/Qualifiers  
 1..3425  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"

gene

CDS

/db\_xref="taxon:10090"  
 /clone="MGC:25823 IMAGE:4165430"  
 /tissue\_type="Salivary gland, 10 week old female mouse"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B"  
 /notes="Vector: pCMV-SPORT6"  
 1..3425  
 /gene="Ssx2ip"  
 /notes="synonym: Adip"  
 /db\_xref="GeneID:99167"  
 /db\_xref="MGI:2139150"  
 240..2087  
 /gene="Ssx2ip"  
 /codon\_start=1  
 /product="synovial sarcoma, X breakpoint 2 interacting protein"  
 /protein\_id="AAH21749.1"  
 /db\_xref="GI:18256806"  
 /db\_xref="GeneID:99167"  
 /db\_xref="MGI:2139150"  
 /translation="MGDMWVTPDVLCTENKNSLYTSETKMSPPSLYSQVLCSSVP LSNVHGFGVCTGENIEQSISYLDQELTTFGPSPSIEESKSKAKRNLINIVAVLNC MNELLVLRKNLLAQSVEQVQKQNIASRATQYNDHVKREYNKRLKRLHOLVNN OLOKNSRLHLLKNEKDEQVQKQNIASRATQYNDHVKREYNKRLKRLHOLVNN KOKNLDVNLNVGRADGKRSWRTDKTEARNEDEMYKILLANDYKROILMENAB LKXVLOOMKEMISLLSPKKPRPRAEDGTVAISDIEDSGELSRDSVMGLSCDT VREQLNSRKWRILKSHVEKLDNOASKVHSEGLNEEDVISRQDHEQTEKLELEIE RCKEMIKAOQQLQOOLATTCDDTTSLRLDYLLEKKEKLEWTLFKQKKNFRR RSFTBAAILGLERKAFEEERASWYKQFLNNFNDFHQNSENKFLFSFSGSDPN LIVHSRPROKHLHSHVANGVPACTSKLTPASPSDTPROTHSCVSEHSSISVLNIT PEEKSPSEARESTDQKSVQSPRSREGCYSCSSAFRSAGHGRDRLP"

ORIGIN

Query Match	99.9%	Score	2688.8	DB 6	Length	3425
Best Local Similarity	99.9%	Pred. No.	0			
Matches	2690	Conservative	0	Mismatches	2	Indels
						Gaps
						0

  

Qy	1	CGTAGAGAGTGCAGAGAGCTGTTAAAGCGTCGACGAGCTGAGCGCGCTCTCCAGGTAT	60
Db	161	CGTAGAGAGTGCAGAGAGCTGTTAAAGCGTCGACGAGCTGAGCGCGCTCTCCAGGTAT	220
Qy	61	CCTGGCTCTGGAACCTTGTATGGAGATGGAGTGTGACAGATGTCCTCCAGGTATGTA	120
Db	221	CCTGGCTCTGGAACCTTGTATGGAGATGGAGTGTGACAGATGTCCTCCAGGTATGTA	280
Qy	121	AGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGTCCTCCAGGTATGTA	180
Db	281	AGAAACAAATCTCTCAATATACCTCAGAAACAAAGATGTCCTCCAGGTATGTA	340
Qy	181	CTCCAGCAGAGTTCGTGCTCTTCACTACCTTATCCAAACAAAGATGTCCTCCAGGTATGTA	240
Db	341	CTCCAGCAGAGTTCGTGCTCTTCACTACCTTATCCAAACAAAGATGTCCTCCAGGTATGTA	400
Qy	241	TGTCTTCTGCAGAGAGAGCAATTTGAAACAAAGATTTCTCTATCTTGTATCAGGAGCTGAC	300
Db	401	TGTCTTCTGCAGAGAGAGCAATTTGAAACAAAGATTTCTCTATCTTGTATCAGGAGCTGAC	460
Qy	301	CACCTTCGGGTTTCCTTCTGTATGAAGATTCCTCAAAAGATGTCCTCCAGGTATGTA	360
Db	461	CACCTTCGGGTTTCCTTCTGTATGAAGATTCCTCAAAAGATGTCCTCCAGGTATGTA	520
Qy	361	AAATATAGTCGCTGTTCTGAACCTGTATGAACGAGCTGCTCGCTTTCAGCGGAGAACCT	420
Db	521	AAATATAGTCGCTGTTCTGAACCTGTATGAACGAGCTGCTCGCTTTCAGCGGAGAACCT	580
Qy	421	GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGAATGACACCACT	480
Db	581	GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGAATGACACCACT	640
Qy	481	GCAGAGCTGCTACGCCAAACTTTAAGAGCAGTTGAAACAGTCCAGCGGAGATGATCGG	540
Db	641	GCAGAGCTGCTACGCCAAACTTTAAGAGCAGTTGAAACAGTCCAGCGGAGATGATCGG	700

**THIS PAGE LEFT BLANK**

Wed Jun 14 13:46:14 2006

[illegible]

## RESULT 4

## RESULT 4

THIS PAGE LEFT BLANK

301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320  
1172 GAAGAGCAGAGGAGCGGACAGGCACTGTTCTATCTCCGATATAGAGATGACTCTGGG 1231  
321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluLeuThr 340  
1232 GAACCTGACGACAGAGCAGCGTGTGGGCGCTTCTGTGACACTGTGACAGAGCAGCTGACA 1291  
341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
1292 AACAGCATCAGGAACAGTGGAGAAATTTTGAAGAGTCAATGATAGAAAACACTCGATACCA 1351  
361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
1352 GCTTCGAAGGTACACTCAGAGGCGCTTATGAGGAGGAGCTCATCTCAGCAGCAAGACCAT 1411  
381 GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla 400  
1412 GAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1471  
401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu 420  
1472 CAGCAGCAGCTTACAGCAGCAGCTGGCCACCACGCTGTGATGATGACACCACTCACTG 1531  
421 LeuArgAspCysThrLeuLeuGluGluLysGluArgLeuLysGluGluThrThrLeuPhe 440  
1532 TTGCGAGAGCTGTACTTCTGCGAGAAAGGAAAGGACCCCTTAAAGAGAGTGGACCTTTT 1591  
441 LysGluGlnLysLysAsnPheGluArgGluArgGserPheThrGluAlaAlaIleArg 460  
1592 AAAGAGCAAAAAAGAAATTTTTCAGAGAGAAAGCGCAAGCTTTACAGAAAGCTGCCATTGCA 1651  
461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480  
1652 TTGGGGTTCGAGAGAAAGCGCTTGAAGAAGAGCGAGCGAGCTGGGTGAAGCAGAGTTT 1711  
481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500  
1712 TTAACATACACCAACTTTGACCAACCAACTCAGAAATGTGAACATTTTCAGTCCCTTC 1771  
501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
1772 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAGCGCCACGCGCAAGAGCTTA 1831  
521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540  
1832 CACAGTGTGGCTAATGGGTGCGAGCTTGCACATCAAACTGACTAAATCTCTTCTGCTGCC 1891  
541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIle 560  
1892 TCACCTTCTACTTTCAGACTTTTCCACAGACACATTCATGTGTCTGACACAGTTCCTCATC 1951  
561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580  
1952 AGTGTCTGATATACTCTGAGAAAGTAAACCAAGTGAAGTTCGAGGAGGAGGAGGAGGAG 2011  
581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600  
2012 GATCAGAAGTGGAGCGTGCAGTGCAGGCCCGAGCTCGCGGAGGGGTGTACAGCGGATGC 2071  
601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
2072 TCCTCGGCGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 2116

RESULT 3  
BC021749  
LOCUS  
DEFINITION  
MUS musculus synovial sarcoma, X breakpoint 2 interacting protein,  
mRNA (cDNA clone MGC:25823 IMAGE:4165430), complete cds.  
ACCESSION  
BC021749  
VERSION  
BC021749.1 GI:18256805  
KEYWORDS  
MGC.

SOURCE  
ORGANISM

Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 3425)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Sapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Schectz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettelman, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Sanchez, A., Whitting, M., Madan, A., Touchman, J.W., Green, E.D.,  
Bouffard, G.G., Blakesley, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26) 16899-16903 (2002)

CONSRTM  
TITLE

Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26) 16899-16903 (2002)

JOURNAL  
PUBMED

Proc. Natl. Acad. Sci. U.S.A. 99 (26) 16899-16903 (2002)

REFERENCE  
AUTHORS

NTH MGC Project  
Direct Submission  
Submitted (18-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: TRAK plate: 30 Row: m Column: 16.  
Location/Qualifiers  
1. 3425  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:25823 IMAGE:4165430"  
/tissue\_type="Salivary gland, 10 week old female mouse"  
/clone\_lib="NCI CGAP\_SG2"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 3425  
/gene="Ssx2ip"  
/note="synonym: Adip"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
240. 2087  
/gene="Ssx2ip"  
/codon\_start=1  
/product="synovial sarcoma, X breakpoint 2 interacting

FEATURES  
source

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: TRAK plate: 30 Row: m Column: 16.  
Location/Qualifiers  
1. 3425  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:25823 IMAGE:4165430"  
/tissue\_type="Salivary gland, 10 week old female mouse"  
/clone\_lib="NCI CGAP\_SG2"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 3425  
/gene="Ssx2ip"  
/note="synonym: Adip"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
240. 2087  
/gene="Ssx2ip"  
/codon\_start=1  
/product="synovial sarcoma, X breakpoint 2 interacting

THIS PAGE LEFT BLANK

sb

protein"		/protein_id="AAH21749.1"		/db_xref="GI:18256806"		/db_xref="GeneID:99167"		/db_xref="MGI:2139150"		/translation="MGDMVTVPVLTCTENKNLSQYTSKMSPLSSLYSQVLCSSVP LSKNVHGVFGVCTGENIEQSIISYLDQELTTFGPSPSLYBSKSKAKREINIVAVLNC MNEELLVQRKNLLAQESVETONLKLSDMDHLOSCYAKLKEQLTSREMLGQLQRDR OLQCKNPSLHQLLKNEDVQKQNIIASRATQYNHVDKREYNKLERLHQLVNN KQDKNTAMDVNTYVRADGRKRSWRTDKTEARNEDEMYKILLANDYEQKQIILMNAE LKKVLOQMKEMISLSPQKKPRERAEDGTVAISDIEDDSGELSDDSVMGLSCDT VREGLTMSIRKQWRLKSHVTKLDNQASKVHSEGLNEEDVISRDQHEOETKLELIE RKCKMIKAQOOLLQOQLATTCDDDTSLRDCVLLBEKRLKEWTLFKQKKNFERE RRSFTEAARLGLERKAFEBERASVWQQLNMTNPHQNSNVKLFSAFSGSSDPDN LIVHSRPROKLLSVANGVPACTSKLTKSPASFTSDPFRQTHSCVSEHSSISVLNIT FEESKPSEVARESTDQKWSVQSRPSRREGCYSGCSSAFRSAHGRDRLP"																																																																																																																																																																																	
Alignment Scores:		Pred. No.:		Length:		3425		Matches:		615		Conservative:		0		Mismatch:		0		Indels:		0		Gaps:		6																																																																																																																																																																	
US-10-644-084-2 (1-615) x BC021749 (1-3425)		QY		1		MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer		20		Db		240		ATGGGAGATTGGATGCTGTGACAGATCCAGTTCTGTGTACAGAAAAA		CAAAAAATCTCTCT		299		QY		21		GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys		40		Db		300		CAATATACCTCAGAAACAAAGATGCTCCGTCAGTTTGTACTCCCGCAGCAAGTTCTGTGC		359		QY		41		SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu		60		Db		360		TCTTCAGTACCTTATCCAAAACGTCGATGGTGTTCGGTGTCTTCGACAGGAGAG		419		QY		61		AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer		80		Db		420		AACATTGAACAAAGTATTTCTCTTGTATCAGGAGCTGACACCTTCGGGTTCCTTC		479		QY		81		LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu		100		Db		480		TTGTATGAAGATCCAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGTCTGTG		539		QY		101		AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal		120		Db		540		AATGTATGAACAGCTGCTGCTGCTTACGCGGAAGAACTTGTGCGCCAGAGAGCGTG		599		QY		121		GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys		140		Db		600		GAGACACAGAACTTGAAGCTGGGACGTGACATGGACCACTGCAGAGCTGTACGCCAAA		659		QY		141		LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg		160		Db		660		CTTAAGAGCAGTTGGAAACGCTCCAGCGGAGATGATCGGCTTCAAGAGAGAGACAGG		719		QY		161		GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal		180		Db		720		CAGCTCAGTGCAGAACAGGAGTTTGCATCAGCTCTCTGAAAGATGAGAAAGATGAGGTA		779		QY		181		GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg		200		Db		780		CAAAATTTACAAATATATATAGCCAGCGGGCTACTCAGTATATATCATGTGTGAAGAGG		839		QY		201		LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys		220		Db		840		AAGGAGCGTGAATATATAAGCTTAAAGAGCGCGCTGCATCAGCTCGTTATGAACAGAG		899		QY		221		AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly		240	

**THIS PAGE LEFT BLANK**



Db 1980 GATCAGAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGTGTCTACAGCGGATGC 2039  
 Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
 Db 2040 TCCTCGGCTTTCAGGAGCGCTCACGGGACCGAGATGATGATCTACCT 2084

RESULT 4  
 BC031527 3410 bp mRNA linear ROD 29-JUN-2004  
 LOCUS Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,  
 DEFINITION mRNA (cdna clone MGC:28268 IMAGE:4008624), complete cds.

ACCESSION BC031527  
 VERSION BC031527.1 GI:21594536  
 KEYWORDS MGC.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 3410)  
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshikiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bonas, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S.G., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Heiton, E., Kettelman, M., Madan, A., Young, A.C., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Touchman, J.W., Green, E.D., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Schmutz, J., Myers, R.M., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Skalska, U., Smalys, D.E., Butlerfield, Y.S., Krzywinski, M.I., Skalska, U., Smalys, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 PUBMED 12477932

2 (bases 1 to 3410)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Speed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 36 Row: k Column: 10  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20270272.  
 Location/Qualifiers  
 1..3410  
 /organism="Mus musculus"  
 /mol\_type="mRNA"

FEATURES  
 source

gene

CDS

# ORIGIN

Alignment Scores:  
 Pred. No.: 0 Length: 3410  
 Score: 3162.00 Matches: 614  
 Percent Similarity: 100.0% Conservative: 1  
 Best Local Similarity: 99.8% Mismatches: 0  
 Query Match: 99.9% Indels: 0  
 DB: Gaps: 0

US-10-644-084-2 (1-615) x BC031527 (1-3410)

Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	248	ATGGGAGATTGGATGACTGTGACAGATCCAGTCTCTGTGTACAGAAACAAAATCTCTCT	307
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
Db	308	CAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTACTCCACGAGATTCTGTGC	367
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyThrGlyGlu	60
Db	368	TCITTCAGTACCTTTATCCAGAAATGTGATGTTTTCGGTGTCTTCTTCACAGAGAG	427
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrPheGlyPheProSer	80
Db	428	AACATTGAACAAAGATTTCTCTATTCATCAGAGCTGACCACTTCGGGTTCCTTCC	487
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu	100
Db	488	TTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAATAATATATAGTCCTCTCTG	547
Qy	101	AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal	120
Db	548	AACGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGAACCTGCTGCGCCACGAGAGCGTG	607
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140
Db	608	GAGACACAGAACTTGAAGCTGGGCGAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA	667

/strain="Czech II"  
 /db\_xref="taxon:10090"  
 /clone="MGC:28268 IMAGE:4008624"  
 /tissue\_type="Mammary tumor metastasized to lung."  
 MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR enhancer."  
 /clone\_lib="NCI CGAP\_Lu30"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 1..3410  
 /gene="Ssx2ip"  
 /note="synonym: Adip"  
 /db\_xref="GeneID:99167"  
 /db\_xref="MGI:2139150"  
 248..2095  
 /gene="Ssx2ip"  
 /codon\_start=1  
 /product="synovial sarcoma, X breakpoint 2 interacting protein"  
 /protein\_id="AAH31527.1"  
 /db\_xref="GI:21594537"  
 /db\_xref="MGI:2139150"  
 /translation="MGDMVTVDVPLCTENKLSQYTSKMSPSLSYQQVLCSSVP  
 LSRNVGVGFCFTEQISYLDQLTFTGFPSPSLYEESKEAKELNIVAVLNC  
 NELLVLRQKLLAQESVETQNLKLGSDMDHLSQYAKLKEQLTSRREMIGLQDR  
 QLOKNSRLHOLLNEDEVKQNI IASRATQYNDHVKRERKYLKRLQHLVWN  
 KQKNLADLVNTYGRADKGGKWRDXTARNEDMYKILLNDYRQKILMENAR  
 LKKVQKMKEMISLLSPQKKPRERADGTGTVASIDTDDSGSLSRVWGLSCDT  
 VREQLTNSIRKQWRLKSHVEKLDNQASKVHSEGLNEEDVSRQDHEQTEKLEIE  
 RCKEMIAQQLLOQLATTCDDTTSLRDCYLLEEKLEWTLPEQKKNFPERE  
 RRSFTAAIRLGLERKAPFEERASWVKQFLNMTNFDHONSENVKLFSPSSGSDPDN  
 LIVHSRPROKHLHVGAVPACTSKLTSLPASPTSDPQTHSCVSEHSSISLVNIT  
 PESKPSVARESTDTQKWSVQSPRSREGSCSAPRSRSHGDRDRLP"

THIS PAGE LEFT BLANK

6a

OS Homo sapiens.  
XX EP1308459-A2.  
XX 07-MAY-2003.  
XX 28-MAR-2002; 2002EP-00007401.  
XX 05-NOV-2001; 2001JP-00379298.  
XX 25-JAN-2002; 2002US-00350978.  
XX (HELI-) HELIX RES INST.  
XX (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX Isegai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX WPI; 2003-450961/43.  
XX N-PSDB; ADB63343.  
XX New polynucleotides and polypeptides, useful for developing a diagnostic  
XX marker or medicines for regulation of their expression and activity, or  
XX as targets of gene therapy.  
XX Claim 1; Page; 222pp; English.  
XX The invention discloses a polynucleotide comprising a sequence selected  
XX from 1970 fully defined nucleotide sequences which encode novel  
XX polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
XX or its partial peptide, an antibody binding to the polypeptide or peptide  
XX of the polynucleotide, immunologically assaying the polypeptide or  
XX peptide of the polynucleotide by contacting the polypeptide or peptide  
XX with the antibody of the encoded protein, and observing the binding  
XX between the two, a transformant carrying the polynucleotide in an  
XX expressible manner and an antisense polynucleotide. The oligonucleotide  
XX is useful as a primer for synthesizing the polynucleotide, or as a probe  
XX for detecting the polynucleotide. The polynucleotides and encoded  
XX proteins are useful as pharmaceutical agents and many disease-related  
XX genes may be included in them, for developing a diagnostic marker or  
XX medicines for regulation of their expression and activity, or as targets  
XX of gene therapy. The genes are involved in tissue and/or cell  
XX regeneration. Membrane proteins, signal transduction-related proteins,  
XX transcription-related proteins, disease-related proteins and genes  
XX encoding them can be used as indicators for diseases (e.g. osteoporosis,  
XX neurological diseases, cancer, tumours). The cDNA may be used to regulate  
XX the activity or expression of the encoded protein to treat diseases. The  
XX sequence presented is a protein of the invention. Note: Some of the  
XX sequence data for this patent is not represented in the printed  
XX specification, but is based on sequence information supplied by the  
XX European Patent Office.  
XX Sequence 504 AA;  
SQ Query Match 68.0%; Score 2152.5; DB 7; Length 504;  
Best Local Similarity 90.3%; Pred. No. 1.6e-143;  
Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;  
QY 1 MGDWMTVPVLCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNKHVGVFCTGE 60  
DB 1 MGDWMTVPDP----ESKTSQYTSKMPSSLYSQVLCSSVPLSKNKHVGFCTED 56  
QY 61 NTEQSYLDQELTTFGFFSLYEEKSKAKRELNTVAVLNCNELLVLQRNLLAQS 120  
DB 57 NTEQSYLDQELTTFGFFSLYEEKSKAKRELNTVAVLNCNELLVLQRNLLAQS 116  
QY 121 ETQNLKPGSDMDHLQSCYAKLEQLETSREMITGLQERDRQLQCKNRLHLLKNEKDEV 180  
DB 117 ETQNLKPGSDMDHLQSCYAKLEQLETSREMITGLQERDRQLQCKNRLHLLKNEKDEV 176  
QY 181 OKLQNIASRATQYNDHVRKREYNKLERHQLVWKKDKNIAMDVINYVGRADKRG 240  
DB 177 OKLQNIASRATQYNDHVRKREYNKLERHQLVWKKDKNIAMDVINYVGRADKRG 236

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 300  
DB 237 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPR 296  
QY 301 ERAEDGTGVAISDIEDSGELSRDVSGLSCDTVRBQLTNSIRKQWRILKSHVEKLDNQ 360  
DB 297 ERVDDSTGTV-ISDVEEDAGELSRDVSGLSCDTVRBQLTNSIRKQWRILKSHVEKLDNQ 355  
QY 361 ASKVHSEGLNEEDVTSRODHEQTEKLELEIERCKEMIKAOQOLLOQALATTCDDDTTSL 420  
DB 356 VSKVHLGPNFDEVTISRODHEQTEKLELEIOCKEMIKAOQOLLOQALATTCDDDTTSL 415  
QY 421 LRDCCVLLERKLEKBEWTLFQKQKNFERERSFTEAAILGLIE 464  
DB 416 LRDCCVLLERKLEKBEWTLFQKQKNFERERSFTEAAILGLIE 459  
RESULT 5  
AAB93250  
XX AAB93250 standard; protein; 417 AA.  
XX AC AAB93250;  
XX DT 26-JUN-2001 (first entry)  
XX DE Human protein sequence SEQ ID NO:12265.  
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX OS Hgmo. sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-00116126.  
XX PR 29-JUL-1999; 99JP-00248036.  
XX PR 27-AUG-1999; 99JP-00300253.  
XX PR 11-JAN-2000; 2000JP-00118776.  
XX PR 02-MAY-2000; 2000JP-00183767.  
XX PR 09-JUN-2000; 2000JP-00241899.  
XX PA (HELI-) HELIX RES INST.  
XX PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.  
XX PS Claim 8; SEQ ID NO 12265; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

THIS PAGE LEFT BLANK

CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
CC present invention  
XX  
XX Sequence 417 AA;  
SQ

Query Match	58.6%	Score	1855.5	DB 4	Length	417			
Best Local Similarity	85.6%	Pred. No.	1.3e-122						
Matches	358	Conservative	27	Mismatches	122	Indels	1	Gaps	1

  

Qy	198	V	K	R	E	R	E	Y	N	K	L	K	E	R	H	O	L	V	M	N	K	D	K	N	T	A	M	D	V	L	N	V	G	R	A	D	K	R	G	S	W	R	T	K	T	E	A	R	N	E	D	E	M	Y	K	257				
Db	1	M	K	R	E	Y	N	K	L	K	E	R	H	O	L	V	M	N	K	D	K	N	T	A	M	D	V	L	N	V	G	R	A	D	K	R	G	S	W	R	T	K	T	E	A	R	N	E	D	E	M	Y	K	60						
Qy	258	I	L	L	N	D	E	Y	R	O	K	I	L	M	E	N	A	E	L	K	V	L	Q	M	K	K	E	M	I	S	L	L	S	P	O	K	K	P	R	E	A	R	E	D	T	G	T	V	A	I	S	D	I	E	317					
Db	61	I	L	L	N	D	E	Y	R	O	K	I	L	M	E	N	A	E	L	K	V	L	Q	M	K	K	E	M	I	S	L	L	S	P	O	K	K	P	R	E	V	D	D	T	G	T	V	-	I	S	D	V	E	119						
Qy	318	D	S	G	E	L	S	R	D	S	V	M	G	L	S	C	D	T	V	R	E	O	L	T	N	S	I	R	K	O	W	R	I	L	K	S	H	V	E	K	L	D	N	O	A	K	S	V	H	S	G	L	E	N	E	D	V	I	S	377
Db	120	D	A	E	L	S	R	E	S	H	W	L	S	C	E	T	V	R	E	O	L	T	N	S	I	R	K	O	W	R	I	L	K	S	H	V	E	K	L	D	N	O	V	K	S	V	H	L	E	G	F	N	D	E	V	I	S	179		
Qy	378	Q	D	E	H	E	T	E	K	L	E	I	E	R	C	K	E	M	I	K	A	Q	Q	L	L	Q	L	A	T	T	C	D	D	T	T	S	L	R	D	C	L	L	E	E	K	E	R	L	K	E	W	437								
Db	180	Q	D	E	H	E	T	E	K	L	E	I	E	R	C	K	E	M	I	K	A	Q	Q	L	L	Q	L	A	T	A	Y	D	D	T	T	S	L	R	D	C	L	L	E	E	K	E	R	L	K	E	W	239								
Qy	438	T	L	P	E	O	K	K	N	F	E	R	R	S	T	E	A	A	I	R	L	G	L	R	K	A	F	E	E	R	A	S	W	V	Q	O	F	L	N	T	N	F	O	H	S	E	N	V	K	L	F	497								
Db	240	S	L	F	E	O	K	K	N	F	E	R	R	S	T	E	A	A	I	R	L	G	L	R	M	A	F	E	E	R	A	S	W	L	K	Q	O	F	L	N	T	T	F	O	H	S	E	N	V	K	L	F	299							
Qy	498	S	A	F	S	G	S	S	D	P	N	L	I	V	H	S	R	P	R	O	K	L	H	S	V	A	N	G	P	A	C	T	S	K	L	T	S	L	P	A	S	P	S	T	S	D	P	R	O	T	H	S	C	V	S	E	H	557		
Db	300	S	A	F	S	G	S	S	D	N	L	I	V	H	S	R	P	O	K	P	H	S	V	N	G	S	P	V	C	M	S	K	L	T	S	L	P	A	S	P	S	T	S	D	F	C	O	T	R	S	C	I	S	E	H	359				
Qy	558	S	S	I	S	V	L	N	I	T	P	E	E	S	K	P	E	V	A	R	E	S	T	O	K	W	S	O	S	R	P	S	R	E	G	C	V	S	G	G	S	A	F	R	S	A	H	G	D	R	D	L	P	615						
Db	360	S	S	I	N	V	L	N	I	T	A	B	E	K	P	Q																																												

RESULT 6  
AAG03000  
ID AAG03000 standard; protein; 137 AA.

XX	AAG03000;
AC	
XX	06-OCT-2000 (first entry)
DT	
XX	Human secreted protein, SEQ ID NO: 7081.
DE	
XX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW	gene therapy; chromosome mapping.
XX	

OS Homo sapiens.  
XX  
XX  
PN EP1033401-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-00200610.  
XX  
PR 26-FEB-1999; 99US-0122487P.  
XX  
PA (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;  
PI  
DR WPI; 2000-500381/45.  
DR N-PSDB; AAC03006.

DR WPI; 2000-500381/45.  
DR N-PSDB; AAC03006.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 FT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 TT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 XX  
 XX  
 PS Claim 13; SEQ ID NO 7081; 71pp + Sequence Listing; English.  
 CC  
 CC The present sequence is a polypeptide encoded by one of a large number of  
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
 CC prepared from total human RNAs or poly(A<sup>+</sup> RNAs derived from 30 different  
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
 CC sequences derived from the 5' ends of mRNAs and even in those cases where  
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
 CC are also used in diagnostic, forensic, gene therapy and chromosome  
 CC mapping procedures. They are used to obtain upstream regulatory sequences  
 CC and to design expression and secretion vectors  
 XX  
 SQ Sequence 137 AA:

Query Match		19.4%;	Score 612.5;	DB 3;	Length 137;
Best Local Similarity		87.0%;	Pred. No. 2.4e-35;		
Matches 120;	Conservative 10;	Mismatches 7;	Indels 1;	Gaps 1;	
Qy	272	MENAEELKVLQQMKEMISLLSPQKKPKPRERADGTGTVASIEDDSGLSDSVDWGLS	331		
Dd	1	MENAEELKVLQQMKEMISLLSPQKKPKPRVDDSTGTIV-ISVVERDAGLSRESMWDL S	59		
Qy	332	CUTVRQLTNSIRKOWRIILSHVKELNDQASKHVSRLNEEDVISRQDHQEOTEKLELI	391		
Dd	60	CETVRQLTNSIRKOWRIILSHVKELDNQSVKHLEGFNDVEDVISRQDHQEOTEKLELI	119		
Qy	392	ERCCKEIMAKOQOLLOQOL	409		
Dd	120	OQCCKEMIKTQOOLLQOOL	137		

RESULT 7  
ADX91485

ID ADX91485 standard; protein; 443 AA.

AC ADX91485;

DT 21-APR-2005 (first entry)

DE Plant full length insert polypeptide seqid 54149.

KW	plant protectant; plant growth regulant; gene therapy; plant;
KW	recombinant DNA construct; physical array; plant breeding marker;
KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW	extreme osmotic condition; pathogen tolerance; pest tolerance;
KW	growth rate; cell cycle pathway; disease resistance;
KW	galactomannan production; lignin production; plant growth regulator;
KW	yield; plant growth; plant development; seed oil; protein yield;
KW	protein content.

Unidentified.

AA  
PN  
US2004034888-A1.XX  
PD  
19-FEB-2004.

28-APR-2003; 2003US-00425114.

AA  
PR 06-MAY-1999; 99US-00304517.

PR 05-NOV-2001; 2001US-00985678.  
XX

PA (LIU//) LIU J.  
PA (ZHOU//) ZHOU Y.

PA (KOV A/) KOVALIC D K.

FA (SCREEN / SCREEN S B.

THIS PAGE LEFT BLANK

26

Db	1007	GAACCTGT---TATTTCCGATGTTGAAGAAGATGCGGGGAACCTAAGCAGACAGAGTATGT	1063
Qy	1062	GGGGCTTTCTCTGTGACACTGTGAGAGACAGCTGACAAAACAGCANTCAGGAAAACAGTGG	1121
Db	1064	GGGACCTTTCTCTGTGAAACTGTGAGAGAGCAGCTTTACAAAACAGCATCAGAAAACAGTGG	1123
Qy	1122	GAATTTGAAAAGTCATGTAGAAAACCTCGATAACCAAGCTTCGAAAGCTACACTCAGAGG	1181
Db	1124	GAATTTTGAAGAAGTCATGTAGAAAAGCTTGATAACCAAGTTTCAAAAGGTACACCTGGAG	1183
Qy	1182	GCCTTAATGAGGAGCAGCTCATCTCCAGCAAGACCATGAGCAAGAGACTGAGAAAACCTGG	1241
Db	1184	GTTTTAATGATGAAAGATGTAATCTCAGCAGACAGCCATGAACTTAAAGTCTCAGCAGC	1243
Qy	1242	AGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGC	1301
Db	1244	AGTTAGAAATTCAGCAGTGTAAAGAAATGATTTAAAACTCAGCAACAGCTTTTACAGCAGC	1303
Qy	1302	AGCTGGCCACACAGTGTGATGATGACACACACTCACTGTTGCGAGACTGTACTTCTGCTGG	1361
Db	1304	AGCTCGCTACTGCTATGATGATGATACCACTTCCTATTAACAGACACTGTTATTTGTTGG	1363
Qy	1362	AGAAAAGGAAGCCCTTAAAGAAAGAGTGGACCCCTTTTAAAGAGCAAAAAGAAATTTTG	1421
Db	1364	AGAAAAGGAAGCTCTCAAGAAAGATGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTG	1423
Qy	1422	AGAGAGAAAGCGGAAGCTTTACAGAGCTGCCATTCGATGGGGTTGGAGAGAAAGCGGT	1481
Db	1424	AGAGGAGAGACGAAGCTTTACAGAAAGCCGTATTCGCTGGGATTTGGAGAGAAAGCAT	1483
Qy	1482	TTGAAGAGAGCGAGCCAGCTGGTAAAGCAGCAGTGTAAACATGACCAACCTTTTGACC	1541
Db	1484	TTGAAGAGAGAGAGCGCAGTTGGTTAAAGCAGCAGTGTCTAATAATGACTTACCTTGACC	1543
Qy	1542	ACCAGAACTCAGAAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACA	1601
Db	1544	ACCAGAACTCAGAAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTCTGATGGGACA	1603
Qy	1602	ATCTTATAGTCCACTCAGCGGCACGGCAAGAGACTACACAGTGCCTGCTTAATGGGGTGC	1661
Db	1604	ATCTTATAGTGCATCTCGAGGCGCGCAAGAGAGCCTCACAGTGTGCTTAATGGGCTC	1663
Qy	1662	CAGCTTGCACATCAAACTGACTAAATCTCTTCTGCTCCTCACTTCTACTTCAGACTTTC	1721
Db	1664	CAGTTTGCATGCTCAAACTTAAATCTCTTCTGCTTCACTTCCACTTCCAGACTTTT	1723
Qy	1722	GCCAGACACATTCATGTGTGCTGAAACAAGTTCATCAGTGTGCTGATATAACTCTGTG	1781
Db	1724	GCCAGACAGCTTCTCTGCATATCTGAACATAGTTTCAATCAATGACTGAATATACTGCTG	1783
Qy	1782	AAGAAGTAAACCAAGTGAAGTGTGCAAGAGAAAGCAGGATCAGAAAGTGGAGCGTGCAGT	1841
Db	1784	AAGAATTTAAACCAAAATCAGGTGGAGAGAAAGTACAAATCAAAATATGAGTGTGGCGT	1843
Qy	1842	CGAGGCCAGCTCGCGGAGGGGTGCTACAGCGGATGTCTCTCGGCCCTTCAGGAGCGTCTC	1901
Db	1844	CAAGACCTGGATCAGAGGAAAGTGTCTATAGTGGATGCTCTCTTGAGCTACACAAATCTCTC	1903
Qy	1902	ACGGGACCGAGATGACTTACCTTAAATGTGGGGCTGCAAGTGTGCTTCCAGATGTGG	1961
Db	1904	ATGTAGAAAAGATGACTTACCTTTAGACATGTGGACTGGAAATTTTTTTCATTAATGTGTT	1963
Qy	1962	CTAGAGG-----AGTTGACACAGGGTGTAGCATAAAGTCAAGTGCCTC-----TAA	2005
Db	1964	CATCAAGTTTTCATCTAAGTTCAAAACAGGGTGTGCTAATAAGTCAAGTATCTCTAATAA	2023
Qy	2006	CTTAAGATGCTCAGAGTTGTTGTTTGGACTTCGCTGTCTTCCCCCAAAGAGCTGAATG	2065
Db	2024	CTTAAGATGCTCAGAGTTGTTGTTTGGACTTCCCTGTCTTCCCCCAAAGAGTTGAATC	2083
Qy	2066	CT--AAGCTACTTAAAGGATGCAAGCTTTGG-----TTGTGTGTTAGTAAACAGGCCCC	2120

2084	DB	TTAAATCTATTAAAGGATATAAAAGCTTTGGATATGTATTTTATGTAAACAGAGGAATC	2143
2121	QY	TGGCTCTGTGACTCCAGGAATGCAT -GGCCCTTTTGGATGGAAAACAAGAGCGCTCGGAATGA -	2178
2144	DB	TGGTCTCTGTGAATAAAGGAATGTATAGATGTTTGGATGGAAAACAAGCACTAGACTGAG	2203
2179	QY	-TTGCCTCGCCAGGTACCGAGAAGAGCACTTTTAGGCACTGGTTCCTGTGTAACATTAAT	2237
2204	DB	TTTTCCTCTTATAGGTATTAAATATAGCACTTTTAGGAAACTGATTATTGTAAATGTTTAA	2263
2238	QY	ATTGCTCCCAAGTGTGGTGTGGCATTTGNAAGTGTAGCCCTTTACTTTGAAATGTATACTGTAGA	2297
2264	DB	TTTTGTCTCAAAATAGTGTGGCATTTGGAAGTTTAGCCCTTTTACTTTGAAATGTATACTGTAGA	2323
2298	QY	TTTTTAAACAAGCAGGTTCTATATTATATATGTTTGTGTGATTTTGGATTTACTCTTTT	2357
2324	DB	TTTTTAAACAAGCGAGTTCATATTATTATATGTTTGTGTG -GTTTGAATTTAGCTCTTT	2382
2358	QY	CATATGTTTTT-----GTGCTGTACATAAAATATACATGACTATG	2396
2383	DB	CATATGTTTTTAAATAAAGTAAATTTATGTATGTTTGTACATAGATACATGATTAATG	2442
2397	QY	TTAAGAGGCTTTTAAAGTTTAAAACTTCACACCATGCTTGTAGTATAGCAATTTTCATGCCAA	2456
2443	DB	TTAAGAGGCTTTTAAAGTTTAAAGTTTTCACA -CAACCATTAAGTATAGTATTTTCATGCC -A	2500
2457	QY	TTAAAAATGTTTTTCAGTGGCATGGTGTTTTACAGAGG - TTAGAGCACTGCCACATGCACAG	2514
2501	DB	GTAAAAATTTTTTTAGTGGTATTCTGTTTACAGATGTATTAGAGCCATTTGATGCATTTACAT	2560
2515	QY	TTAAGACTTTATTTTTTAAGCCATCTGGGCNAATAAAATTTCAAGGCCCTTTTCAT -AAGCTG	2573
2561	DB	TTAAGA -ATTCTCTTTTAAATACATCTGGGCNAATAAATTTGAAAGGTATTTCCATGAAGCTG	2619
2574	QY	AGTTTCT--AGATAACTAGAACTACTAAACGTTCATTTTTTGAGATTTTTTAAAGCAATT -GTA	2629
2620	DB	AGTTCTTTAGTAAATCAACACTACTTAACATTACATTTTTTGGATTTTTTATGATAGATAGAT	2679
2630	QY	TTTTTATTTTATATATGT -GAATGTTATTAATTTCTAAGAGGAATATTGATTATGGAGTAAT	2688
2680	DB	TTTTTATTTTGTATGTAGAATAATTATAATTTTTTAAAGGAGCTATTGTGATAGAGAAT	2739
2689	QY	GGGG	2692
2740	DB	AGGG	2743

## RESULT 4

AAH14625	AAH14625 standard; cDNA; 2716 BP.
XX	
II	
XX	
XX	AAH14625;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:12264.
XX	
KW	Human; primer; detection; diagnosis;
XX	
OS	Homo sapiens.
XX	
PN	<u>EP1074617-A2.</u>
XX	
PD	07-FEB-2001.
XX	
XX	28-JUL-2000; 2000EP-00116126.
XX	
PR	29-JUL-1999; 99JP-00248036.
PR	27-AUG-1999; 99JP-00300253.
PR	11-JAN-2000; 2000JP-00118776.
PR	02-MAY-2000; 2000JP-00183767.
PR	09-JUN-2000; 2000JP-00241899.
XX	

THIS PAGE LEFT BLANK



(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;  
WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent polynucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;

Query Match 46.8%; Score 1260.4; DB 4; Length 2716;  
Best Local Similarity 79.7%; Pred. No. 0;  
Matches 1716; Conservative 0; Mismatches 376; Indels 60;

594	QY	TCTGGAAGANTGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCCGGCTA	653
	Db	TACTAAAGAANTGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATTTGCAAGTCGAGCTA	60
654	QY	CTCAGTATAATCATGATCTGGAAGGAGGAGCGTGTAATATAATAGCTAAAGGAGGCC	713
61	Db	CTCAGTATAATCATGATATGAGAGAGAAAGAGCGTGTAATATAATAACTGAAGGAACGTC	120
714	QY	TGCATCAGCTCGTTATCGAACAGAAGGATATAAACATAGCCNTGGATGTTTTAAATTATG	773
121	Db	TACATCAACTGTTATGAAACAGAAAGATAGAAAATATAGCTATGGACAATTTTGAATTTATG	180
774	QY	TGGGTGAGCTGATGCGCAAACGAGGCTCATGAGGAGCTGACAAAACAGAAAGCCAGGAATG	833
181	Db	TCGGGAGAGCTGATCGAAAAGAGGCTCTCTGGAGGACTGCTAATAACTGAAGCCAGGATG	240
834	QY	AAGATCAGATGTACAAAATTCCTCTGAAATGATATAGATACCGCAGAGACAGATCCTGA	893
241	Db	AGATGAAATGTATATAAAATTCCTCTGAATGATTTATGAATATCGTCAGAAACAAATCCTAA	300
894	QY	TGGAGAACCGGAGCTGGAAGAAAGGTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCC	953
301	Db	TGGAAAATCGAAGACTTTAGAGGTTCTTCAACAATGAAAAGGAAATGATTTCTCTTC	360
954	QY	TGTTCTCTCAGAAAGAGACCCAGGAAAGAGCAGAGGACGGCAGCAGGCACTGTTGCTA	1013
361	Db	TTTCTCCCCAAAAGAGAAACCTTAGAGAAAGAGTAGATAGTACAGGAACTGT---TA	417
1014	QY	TCCTCCGATATAGAGATGACTCTGGGGAACTGAGCAGACAGCGTGTGGGGCCTTTCTC	1073

Db	418	TTTTCCGATGTTGAAGAAGATGCGGGGAACTAAGCAGAGAGAGATATGTGGGACCTTTTCCT	477
Qy	1074	GTGACACTGTGAGAGAGCAGCTGCACAAAACAGCATCATCAGAAACAGTGGAGAAATTTTGAAA	1133
Db	478	GTGAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGAAA	537
Qy	1134	GTCAATGTAGAAAACTCGATAACCAAGCTTCCAAAGGTTCACAGAGGCGCTTAATGAGG	1193
Db	538	GTCAATGTAGAAAACTGTATACCAAGTTTCAAAAGGTACACCTGGNAGGTTTAAATGATG	597
Qy	1194	AGGAGCTCATCTCAGACACAGACCATGACGAAGAGACGTGAGAAACTGGAGCTGGAGATTTG	1253
Db	598	AAGATGTAACTCTCAGCAAGAACCAAGAACTGAAAAAACTCGAGTTAGAAAATTC	657
Qy	1254	AGCGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACACGACGAGCTGGCCACCA	1313
Db	658	AGCAGTGTAAAGAAATGATTAAGCTCAGCAACAGCTTTTACGACGAGCTCCGCTACTG	717
Qy	1314	CGTGTGATGTGACACCACTCAGCTGTGTTCGAGACTGTATTCTTGGAAAGAAAGGAAC	1373
Db	718	CATATGATGATGATACCACTTCACTATTACGAGACTGTATTCTTGGNAGAAAGGAAC	777
Qy	1374	GCCTTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGGC	1433
Db	778	GTCTCAAGAGAAATGGTCCCTTTTAAAGAGCAGNAAAAAGAAATTTTGAGAGGGAGAGAC	837
Qy	1434	GAAGCTTTACAGAAAGCTGCCATTTCGATTTGGGTTTGGAGAGAAAGGCGTTTGAAGAGAGC	1493
Db	838	GAAGCTTTACAGNAGCCGCTATTTCGCTTGGGATTGGAGAGANTGGCAATTTGAAGAGAAA	897
Qy	1494	GAGCCAGCTGGGTAAAGCAGCAGTTTTTAAACATGACGAACTTTTGACCACCAAGAACTCAG	1553
Db	898	GAGCCAGTTGGTTAAAGCAGCAGTTTTCTAAATATGACTACCTTTTGACCACCAAGAACTCAG	957
Qy	1554	AAAATGTGAAACTTTTCAGTGCCTTCTCAGAAAGTTCTGATCCAGACAATCTTTATAGTCC	1613
Db	958	AAAATGTGAAACTTTTCAGTGCCTTCTCAGAAAGTTCTGATTTGGGACAACTCTTATAGTGC	1017
Qy	1614	ACTCAGCGCCACGGCAAAAGAGCTACACAGTGTGCTAATGGGGTGCAGCTTGCACAT	1673
Db	1018	ACTCGAGCGCCGCAAAAGAGCCTCACAGTGTGCTAATGGGTCTCCAGTTTGCATGT	1077
Qy	1674	CAAACTGACATAACTCTCTTCTGCTCCACTTCTACTTTCAGACTTTTCGCCACAGACATTT	1733
Db	1078	CTAACTTACTAAATCTCTCTGCTTCCACTTCCACTTTCCAGACTTTTGCCACAGACGTTT	1137
Qy	1734	CATGTGTCTGTAACACAGTTTCCATCAGTGTGCTGTAATATACTCTCTGAAGAAAGTAAAC	1793
Db	1138	CTTGATATCTGAACATAGTTTCAATCAATGTACTGAATATACTGTGGAAGAAATTAAC	1197
Qy	1794	CAAGTCAAGTTGCAAGAGAAAGCAGGATCAGAAAGTGGAGCGTGCAGATCGAGGCCAGCT	1853
Db	1198	CAAACTCAGTTGAGAGAGAAATGTACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGAT	1257
Qy	1854	CGCGGAGGGTGCTTACAGCGGATGCTCTCTCGGCTTTCAGGAGCGCTCACGGGGACCGAG	1913
Db	1258	CACAGGAAGGTTGCTATAGTGGATGCTCTCTGAGCTACACAAATCTCTCATGTAGAAAAAG	1317
Qy	1914	ATGACTTACTTAAATGTCGGGGCTCAGTGTGTTCCAGATGTGCGCTAGAGG-----	1968
Db	1318	ATGACTTACTTATAGACATGTGACATGGAAATTTTTTTTCAATTAATGTGTTTCATCAAGTTTCA	1377
Qy	1969	-----AGTTGACACAGGGGTAGCATAAAGTCAGTCCGTC-----TAAGTTAAAGATGCTC	2017
Db	1378	CATCTAAGTTGAAAACAGGGTGTGTCATAAAGTCAGTTATCTCTAATAACTTTAAGATGCTC	1437
Qy	2018	AGAGTTGTTTGTGGAATTCGCTGTCTTCCGCCAAAGAGCTGAAATGCT--AAGCTACTT	2076
Db	1438	TGAGTTGTTTGTGGAATTCGCTGTCTTCCGCCAAAGAGTTGAAATCTTAAATCTAATTT	1497
Qy	2077	AAAAAGATCAAGCTTTGG-----TTGTGTGTGTAGTAAACAGAACCCCTGGCTCTGTGAC	2132
Db	1498	AAAAAGATATAAAGCTTTGGATATGTATTTTTTATGTAACAGAGCATCTGGTCTGTGAA	1557

THIS PAGE LEFT BLANK

2133 TGCAGGAATGCAT-GGCGTTTGGATGGAACAGAGCGCTGGATGA--TTGCTCGCCA 2189  
 1558 TAAAGGAATGTATAGATGTTTGGATGGAACAAACAGCACTAGACTGAGTTTCTCTTATA 1617  
 2190 GGTACCCAGAGAGCACTTTTAGGCACTGGTTCTGTAACACATTAATAATTCGTCCCAAG 2249  
 1618 GGTATTAAATAAGCACTTTTAGGAACTGATTAATGTAATGTTTAAATTTGCTCAA 1677  
 2250 TGTGGTTGGATGGAAGTGTAGCTTTTACCTTGTGAATGTATCTGTAGATTTTAAACAAAG 2309  
 1678 TATAGTTGGCATTGGAAGTTAGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 1737  
 2310 CAGGTTCTATATTATTTATGTTTGTAGTGTGTTTGGGATTAACCTCTTTCATATGTTTT-- 2367  
 1738 CGAGTTCTATATTATTTATGTTTGTAGTGTG-GTTTTGAATTAACCTCTTTCATATGTTTTAA 1796  
 2368 -----GTGTCGTGTACATAAATATACATGACTATGTTAAAGAGGCTTT 2408  
 1797 ATAAAGTGAAATTTATGTATGTTTGTACATAGATACATGATTAATGTTAAAGGCTTT 1856  
 2409 AAGGTTTAAATCTTCAACCATGCTTGTAGTATAGCATTTCAATGCCAATTAATAATGTTTT 2468  
 1857 AAGATTTAAAGTTTCA-CAACCATAGATATAGTATTTTCATGCC-AGTAAATTTTTT 1914  
 2469 CAGTGGCATGTTGTTTACAGAGG--TTAGGACCACTGCCACATCAGATTAAGACTTTAT 2526  
 1915 TAGTGGTATTTCTGTTTACAGATGTTATGAGGACCATGATGATTAACATTTAAAGA-ATTCT 1973  
 2527 TTTTAAGCATCTGGGCAATAAATAATCAAGCCCTTCAT-AAAGTGAGTTC---AGAT 2582  
 1974 CTTTAAATACATCTGGGCAATAAATAATGAAAGGTATTTCCATGAAGCTGAGTCTTTAGAT 2033  
 2583 AACTAGAACTACTAACGTTACATTTTGTAGATTTTAAAGCAATTTTAAAGCAATTTTAA 2641  
 2034 AATCAACACTACTAACATTTTGTAGATTTTGTAGATTTTAAAGCAATTTTAAAGCAATTTTAA 2093  
 2642 TATGT-GAATGTTAATTTTCAAGAGAAATTTGATATGAGTAAATGGG 2692  
 2094 TATGTAGAAATTAATAATTTTAAAGGCACTATTGATGATAGAAATAGGG 2145

## RESULT 5

ADB63343

ID ADB63343 standard; cDNA; 2272 BP.

AC ADB63343;

04-DEC-2003 (first entry)

Human cDNA encoding clone TESTI20071130.

Human; ss; gene; pharmaceutical; diagnostic; gene therapy;

tissue regeneration; cell regeneration; membrane protein;

signal transduction-related protein; transcription-related protein;

osteoporosis; neurological disease; cancer; tumour.

Homo sapiens.

Key Location/Qualifiers

CDS 224..1738

/\*tag= a

/product= "Clone TESTI20071130 protein"

EP1308459-A2.

07-MAY-2003.

28-MAR-2002; 2002EP-00007401.

05-NOV-2001; 2001JP-00379298.

25-JAN-2002; 2002US-00350978.

(HELI-) HELIX RES INST.  
 (REAS-) RES ASSOC BIOTECHNOLOGY.

Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
 Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
 Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
 WPI: 2003-450961/43.  
 P-PSDB; ADB65313.

New polynucleotides and polypeptides, useful for developing a diagnostic  
 marker or medicines for regulation of their expression and activity, or  
 as targets of gene therapy.

Claim 1; Page: 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected  
 from 1970 fully defined nucleotide sequences which encode novel  
 polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
 or its partial peptide, an antibody binding to the polypeptide or peptide  
 of the polynucleotide, immunologically assaying the polypeptide or  
 peptide of the polynucleotide by contacting the polypeptide or peptide  
 with the antibody of the encoded protein, and observing the binding  
 between the two, a transformant carrying the polynucleotide in an  
 expressible manner and an antisense polynucleotide. The oligonucleotide  
 is useful as a primer for synthesizing the polynucleotide, or as a probe  
 for detecting the polynucleotide. The polynucleotide and encoded  
 proteins are useful as pharmaceutical agents and many disease-related  
 genes may be included in them, for developing a diagnostic marker or  
 medicines for regulation of their expression and activity, or as targets  
 of gene therapy. The genes are involved in tissue and/or cell  
 regeneration. Membrane proteins, signal transduction-related proteins,  
 transcription-related proteins, disease-related proteins and genes  
 encoding them can be used as indicators for diseases (e.g. osteoporosis,  
 neurological diseases, cancer, tumours). The cDNA may be used to regulate  
 the activity or expression of the encoded protein to treat diseases. The  
 sequence presented is a cDNA of the invention. Note: Some of the sequence  
 data for this patent is not represented in the printed specification, but  
 is based on sequence information supplied by the European Patent Office.

Sequence 2272 BP; 783 A; 398 C; 502 G; 589 T; 0 U; 0 Other;

Query Match 46.8%; Score 1258.8; DB 10; Length 2272;

Best Local Similarity 77.2%; Pred. No. 0;

Matches 1651; Conservative 0; Mismatches 337; Indels 150; Gaps 4;

Qy	1	CGTAGAGAGTGACAGAGAGCTGTTGTAAGCGTCGACAGCTAGCCGCTCCCTCAGGTAT	60
Db	145	CATAGAAGAGTGACAGAGCTAGACTAATAATGTTTAACTGCTGAACCTAGTTCCTCAGGTAT	204
Qy	61	CCTGGCTCTGGAACCTTGCTATGGAGATTGGATGACTGTGACAGATCCAGTTCCTGTGAC	120
Db	205	CCTGGCTCTGAGATTGCTATGGAGATTGGATGACTGTGACAGATCCAGTTCCTGTGAC	252
Qy	121	AGAAAAACAAAATCTCTCTCAATATACCTCAGAAAAACAAAGATGTCTCCGTCAGATTGTA	180
Db	253	AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAACAAAGATGTCTCCATCAAGTTATA	312
Qy	181	CTCCAGCAAGTTCGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGG	240
Db	313	CTCAGCAAGTTCGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGG	372
Qy	241	TGTCCTCTGCACAGGAGAGAACATTGAACAAAGTATTTCTTATCTGATCAGGAGCTGAC	300
Db	373	TGCTCTCTGCACAGGAGAGAACATTGAACAAAGTATTTCTTATCTGATCAGGAGCTGAC	432
Qy	301	CACCTTCGGGTTTCTCTTCTTGTATGAAGAAATCCAAAGTAAAGAGCAAGAGAGAAATT	360
Db	433	TACTTTTGGTTTCTCTTCTTGTATGAAGAAATCCAAAGTAAAGAGCAAGAGAGAAATT	492
Qy	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTCTGCTGCTCAGCGGAAGAACCT	420
Db	493	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTCTGCTGCTCAGCGGAAGAACCT	552

THIS PAGE LEFT BLANK

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 22:24:24 ; Search time 498 Seconds  
(without alignments)  
10114.508 Million cell updates/sec

Title: US-10-644-084-1  
Perfect score: 2692  
Sequence: 1 cgtcaggagtgacaggagc.....ttgattatgagtaatggggg 2692

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1403666 seqs, 935554401 residues

Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/ina/1 COMB.seq:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/ina/5 COMB.seq:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6A COMB.seq:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/ina/6B COMB.seq:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/ina/7 COMB.seq:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/ina/8 COMB.seq:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/ina/9 COMB.seq:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/ina/PP COMB.seq:\*
- 9: /EMC\_Celerra\_SIDS3/ptodata/2/ina/RE COMB.seq:\*
- 10: /EMC\_Celerra\_SIDS3/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258.8	46.8	2272	3	US-10-104-047-1497
2	297.2	11.0	428	3	US-09-513-999C-3004
3	73.6	2.7	7218	2	US-08-232-463-14
4	58.2	2.2	198	3	US-09-513-999C-29101
5	55.8	2.1	2093	3	US-10-104-047-1666
6	50.8	1.9	1141	3	US-09-806-708B-22
7	48.8	1.8	5883	3	US-09-949-016-5001
8	48.8	1.8	28806	3	US-09-949-016-13217
9	46.6	1.7	1312	2	US-09-976-594-886
10	46	1.7	3489	2	US-08-728-323A-1
11	46	1.7	3489	3	US-09-298-568-1
12	46	1.7	3489	3	US-09-410-379-1
13	46	1.7	3489	3	US-09-894-273-1
14	46	1.7	32207	2	US-08-770-379-20
15	46	1.7	32207	2	US-08-757-669A-20
16	46	1.7	32207	3	US-09-230-371A-20
17	44.4	1.6	543	10	5273901-6
18	44.4	1.6	832	3	US-09-621-976-2813
19	43.6	1.6	16442	3	US-08-781-891-208
20	43.6	1.6	16442	3	US-09-618-166-208
21	43.4	1.6	2357	3	US-10-104-047-1780
22	43.4	1.6	15252	3	US-09-949-016-13584
23	43	1.6	3337	2	US-08-072-610-1

24	43	1.6	3337	2	US-08-719-822B-1	Sequence 1, Appli
25	43	1.6	3337	3	US-09-092-458-1	Sequence 1, Appli
26	43	1.6	3337	3	US-08-719-821C-1	Sequence 1, Appli
27	42.8	1.6	1995	2	US-08-425-069-3	Sequence 3, Appli
28	42.8	1.6	1995	2	US-08-317-844B-3	Sequence 3, Appli
29	42.4	1.6	1852	3	US-09-969-852-4	Sequence 4, Appli
30	42.2	1.6	150394	3	US-09-949-016-13042	Sequence 13042, A
31	42	1.6	1664976	3	US-08-916-421B-1	Sequence 1, Appli
32	42	1.6	1664976	3	US-09-692-570-1	Sequence 1, Appli
33	41.8	1.6	1745	3	US-09-949-016-4711	Sequence 4711, Ap
34	41.8	1.6	5746	3	US-09-949-016-16453	Sequence 16453, A
35	41.6	1.5	289	3	US-09-007-005-17	Sequence 17, Appl
36	41.6	1.5	289	3	US-09-244-796-17	Sequence 17, Appl
37	41.4	1.5	92227	3	US-09-949-016-11929	Sequence 11929, A
38	41.4	1.5	92227	3	US-09-949-016-15421	Sequence 15421, A
39	41.2	1.5	929	3	US-09-671-317-14	Sequence 14, Appl
40	41.2	1.5	1001	3	US-09-671-317-439	Sequence 439, App
41	41	1.5	2338	2	US-08-425-069-1	Sequence 1, Appli
42	41	1.5	2338	2	US-08-317-844B-1	Sequence 1, Appli
43	41	1.5	2532	3	US-09-328-352-1168	Sequence 1168, Ap
44	41	1.5	640681	3	US-09-790-988-1	Sequence 1, Appli
45	40.8	1.5	330	4	US-10-211-689-123	Sequence 123, App

ALIGNMENTS

RESULT 1  
US-10-104-047-1497  
; Sequence 1497, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cdna  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1497  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1497

Query Match	46.8%;	Score 1258.8;	DB 3;	Length 2272;
Best Local Similarity	77.2%;	Pred. No. 0;		
Matches 1651;	Conservative	0;	Mismatches 337;	Indels 150; Gaps 4;
QY	1	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGACGACCTGAGCCGCGCTCTCCAGGTAT	60	
Db	145	CATAGAAGAGTGACAGCAGCTAGACTAAATGTTAACTGCTGAACCTAGTCTCCTCAGGTAT	204	
QY	61	CCTGGCTCTGGAACTTGCTATTGGGAGATTGCGATGACCTGTGACAGATCCAGTTCGTGTAC	120	
Db	205	CCTGGCTCTAGAGATTGCTATTGGGAGATTGCGATGACTGTTACAGATC-----C	252	
QY	121	AGAAACAAAAATCTCTCTCAATATACCTCAGAAACAAAGATGTCCTCGGCCAGTTTGTA	180	
Db	253	AGAAAGCAAACTATCTCTCAATATACCTCAGAAACAAAGATGTCCTCAAGTTTATA	312	
QY	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGCATGGTGTTCGG	240	
Db	313	CTCAGCAAGTGCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	372	
QY	241	TGCTCTTCGCACAGGAGAAACATTGAACAAAGTATTTCTATCTTGTATCAGGAGCTGAC	300	
Db	373	TGCTCTTCGCACAGAAATATATTGAACAGAGTATCTCATATCTTGTATCAGGAATTGAC	432	
QY	301	CACCTTCGGGTTTCCTCTTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAAATT	360	

**'THIS PAGE LEFT BLANK'**

Db 433 TACTTTTGGTTTCTCTTATATATGAAGATCAAGATTAAGAGACAAAGAGAGAGTT 492  
Qy 361 AATATAGTCGCTGTTCTGAACTGTATGACAGCTGCTCGTCTTACGCGAAGAACTT 420  
Db 493 AATATAGTAGCTGTACTAAATTTGTATGAATGAGTGTCTTGTCTTACGCGAAGAACTT 552  
Qy 421 GCTGCGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTTGGCGAGTGACATGGACCACT 480  
Db 553 TCTAGCTCAGGAAATGTGGAGACACAGAAATTTGAAGCGGGAAGTGTATGACCACTT 612  
Qy 481 CGAGAGCTCTACGCGAACTTTAAGAGCAGTTTGAAGAGCTCCAGCGGAGATGATCGG 540  
Db 613 ACAGAGCTCTACTCAAACTTTAAGGAACAACCTGGAACCTCCAGAGGGAATGATGG 672  
Qy 541 GCTTCAAGAGAGACAGCAGCTGCGAGTGCAGACAGAGAGTGTGACATCAGCTCTGAA 600  
Db 673 GCTTCAGGAAGAGACAGACGTTTCAATGTGAAGACAGGAATTTGCACTCAGCTACTAAA 732  
Qy 601 GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTTACTCAGTA 660  
Db 733 GAATGAGAAAGATGAGGTGCANAATTTACAAAATATCATTTGCAAGTCCGAGCTACTCAGTA 792  
Qy 661 TAATCATGATGTGAAGAGAGAGCGTGAATATTAATAGCTTAAAGAGCGCTCTCATCA 720  
Db 793 TAATCATGATGTGAAGAGAGAGCGTGAATATTAATAGCTTAAAGAGCGCTCTCATCA 852  
Qy 721 GCTCGTTATGAACAGAGAGTAAACACATAGCCATGGATGTTTAAATTTATGTGGTGG 780  
Db 853 ACTTGTGTATGAACAAGAGATGAAGAAATAGCTATGGACATTTTGAATTTATGTGGGAG 912  
Qy 781 AGCTGATGCAACAGAGGCTCATGAGGACTGACAAAACAGAAAGCCAGGAATGAAGATGA 840  
Db 913 AGCTGATGCAACAGAGGCTCTCGAGGACTGTTAAACTGAAGCCAGGAATGAAGATGA 972  
Qy 841 GATGTACAAAATTTGTTGAATGATTAAGATACCGCCAGAGAGAGATCTCTGATGGAGAA 900  
Db 973 AATGTATAAAATTTCTTGAATGATTAAGATATCGTCAGAAAACAAATCTTAATGGAANA 1032  
Qy 901 CCGGAGCTGAGAGGTTCTCCAGCAGATGAGAGAGGATGATCTCTCTCTCTCTCC 960  
Db 1033 TCGAAGAACTTAAGAGGTTCTTCAACAAATGAAGAAAGGAATGATTTCTCTCTCTCTCC 1092  
Qy 961 TCAGAGAGAAAGCCAGGGAAGAGCAGAGAGCGCACAGGCACTGTTGCTATCTCCGA 1020  
Db 1093 CCAAGAGAAACCTTAGAAGAGAGTAGATGATAGTACAGAACTGT---TATTTCCGA 1149  
Qy 1021 TATAGAGATGACTCTGGGGAATCTGAGCAGACAGCGTGTGGGCGCTTTCTCTGACAC 1080  
Db 1150 TGTTCAGAGAAAGATGCGGGGAACCTAAGCAGAGAGAGTATGTGGGACCTTTCTCTGTGAAC 1209  
Qy 1081 TGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTGAAAGTCATGT 1140  
Db 1210 TGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTGAAAGAGTCATGT 1269  
Qy 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGACGT 1200  
Db 1270 AGAAAGCTGTATAACCAAGTTTCAAGGTACACCTGAGGTTTATGATGAGAGATGT 1329  
Qy 1201 CATCTCAGCACAAGACCATGACACAGAGACTGAGAACTGAGAGCTGGAGATTGAGCGGTG 1260  
Db 1330 AATCTCAGCACAAGACCATGACACAGAACTGAAAACTCGAGTTAGAAAAATTCAGCAGTG 1389  
Qy 1261 TAAAGAGATGATCAAGCTCAGCAGCAGCTTACAGCAGCAGCTGCGCACCAACAGTGA 1320  
Db 1390 TAAAGAAATGATTAACCACTCAGCAACAGCTTTTACAGCAGCAGCTGCGTACTGCAATGA 1449  
Qy 1321 TGATGACACCACTCACTCTTTCGAGACTGTTTACTTGTGGAAGAAAAAGAAAGCGCTTAA 1380  
Db 1450 TGATGATACCACTTCACTATTACGAGACTGTTATTGTTGGAAGAAAAAGGACGCTCAA 1509  
Qy 1381 AGAAGAGTGAACCTTTTAAAGAGCAAAAGAAATTTTGAAGAGAAAGCGGAGCTT 1440  
Db 1510 AGAAGAAATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAAGGGGAGAGACGAAGCTT 1569

Qy 1441 TACAGAGCTGCCATTCGATTGGGTTGGAG----- 1471  
Db 1570 TACAGAGCCGCTATTCGCTGGGATTGGAGATTGGCTTTCTCTGCAAGTGTTTTACA 1629  
Qy 1472 ----- 1471  
Db 1630 CAAGAAATCTCTTTTGGGTATCCATCATTCATCCAGCTCAGTAGGTGTGTGTGTCAGAAG 1689  
Qy 1472 -----AGAAAGCGTTTGAAGAGCGAG 1496.  
Db 1690 ACAGTGAACCAAGACTATATTGCTGTTTAAAAAAGAAAGGCAATTTGAAGAGAAAGAG 1749  
Qy 1497 CCAGCTGSGTAAAGCAGCAGTTTTTAAACATGACGAACCTTTGACACACAGAACTCAGAAA 1556  
Db 1750 CCAGTTGTTAAAGCAGCAGTTTTCTAAATATGACTACTTTTGACACACAGAACTCAGAAA 1809  
Qy 1557 ATGTGAAACTTTTTCAGTGCCTTCTCAGAAAGTTCTGATCCAGACAAATCTTATAGTCCACT 1616  
Db 1810 ATGTGAAACTTTTTCAGTGCCTTCTCAGAAAGTTCTGATTTGGGACAAATCTTATAGTCCACT 1869  
Qy 1617 CAGGCGCAGCGCAAAAGAGCTACACAGTGTGGCTAATGGGTGCCAGCTTGCACATCAA 1676  
Db 1870 CGAGGCGAGCGCAAAAGAGCGCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTA 1929  
Qy 1677 AACTGACTTAAATCTTCTTCCCTCAGCTTCTACTTCTCAGACTTTTCCGCGACACATTCAT 1736  
Db 1930 AACTTACTTAAATCTTCTTCCCTCAGCTTCTTCCCTCAGACTTTTCCGCGACACATTCAT 1989  
Qy 1737 GTGTGTCTGAACACAGTTCCATCAGTGTCTGATATAAATCTCTGAAGAAAGTAAACCAA 1796  
Db 1990 GCATATCTGAACATAGTTCAATCAATGATCTGATATAAATCTGCTGAAGAAATTAACCAA 2049  
Qy 1797 GTGAGTTGCAAGAAAGACCGATCAGAGTGGAGCGTGCAGTCCAGGCGCCAGCTCGC 1856  
Db 2050 ATCAGGTTGGAGGAGAACGTCACAAATCAAAATGAGAGTGGCGTCAAGACCTGGATCAC 2109  
Qy 1857 GGGAGGCGTGTCTCAGCGGATGCTCTCGGCTTCCAGAGCGCTCACGGGACCGAGATG 1916  
Db 2110 AGGAGGTTGCTATGATGAGTGTCTCTTGGAGTACACAAATCTCATGTAGAAAAAGATG 2169  
Qy 1917 ACTTACCTTAAATGTGCGGCTGCAGTGTCTTCCAGATGTGCGCTAGAGG----- 1968  
Db 2170 ACTTACCTTGAACATGCGACTGGAAATTTTTCATTAATGTGTTCATCAAGTTTTCAT 2229  
Qy 1969 ---AGTTGACACAGGTTGATAGCATAAAGTCAGTCTCT 2003  
Db 2230 CTAAGTTGAAACAGGCGTGTGTATAAAGTCAGTTTATCT 2267

## RESULT 2

US-09-513-999C-3004  
; Sequence 3004, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.Y.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3004  
; LENGTH: 428  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS

NO PAGE LEFT BLANK



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:34 ; Search time 51 Seconds  
(without alignments)  
1055.516 Million cell updates/sec

Title: US-10-644-084-2  
Perfect score: 3165  
Sequence: 1 MGDWMTVDVLTCTENKNLS.....CYSCSSAFSAHGRDDLP 615

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pap.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pap.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pap.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pap.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pap.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pap.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2152.5	68.0	504	US-10-104-047-3467
2	612.5	19.4	137	US-09-513-999C-7081
3	223.5	7.1	3259	US-09-949-016-6507
4	222.5	7.0	905	US-09-248-796A-16333
5	222.5	7.0	2733	US-09-949-016-11433
6	220.5	7.0	1780	US-09-949-016-6899
7	220.5	7.0	1786	US-09-949-016-7880
8	215.5	6.8	2125	US-09-913-172-29
9	212.5	6.7	2663	US-09-538-092-1252
10	211.5	6.7	1388	US-08-685-576-1
11	211.5	6.7	2662	US-09-595-684B-31
12	210.5	6.7	1354	US-08-685-871-2
13	208.5	6.6	1388	US-08-685-576-4
14	208.5	6.6	1388	US-09-976-594-296
15	208	6.6	2704	US-09-538-092-1260
16	206.5	6.5	1676	US-09-949-016-7610
17	206.5	6.5	2186	US-09-949-016-10828
18	206.5	6.5	2349	US-09-538-092-914
19	204	6.4	1427	US-09-538-092-1044
20	203	6.4	3210	US-09-538-092-1154
21	203	6.4	3248	US-08-353-700-1
22	203	6.4	3248	PCT-US95-16216-1
23	200.5	6.3	1958	US-10-028-946-4
24	200.5	6.3	2053	US-09-964-956-11
25	200.5	6.3	2054	US-10-028-946-2
26	200.5	6.3	2066	US-09-964-956-9

ALIGNMENTS

RESULT 1  
US-10-104-047-3467  
; Sequence 3467, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cdna  
; FILE REFERENCE: HI-A0105  
; CURRENT APPLICATION NUMBER: US/10/104.047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3467  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3467

Query Match	68.0%;	Score	2152.5;	DB 2;	Length	504;			
Best Local Similarity	90.3%;	Pred. No.	9.2e-157;						
Matches	419;	Conservative	20;	Mismatches	20;	Indels	5;	Gaps	2;
Qy	1	MGDMWTTDVPVLTCTENKNLSQYTS	ETKNSPSSLYSQOVLCS	SVPLSKNVHGVFGVCTGE	60				
Db	1	MGDMWTTDVPVLTCTENKNLSQYTS	ETKNSPSSLYSQOVLCS	SVPLSKNVHGVFGVCTGE	56				
Qy	61	NIEQISYILDQELTTFGFPSPSLYEESKRE	AKELNIVAVLNCMKNELLVLRK	NLLAQESV	120				
Db	57	NIEQISYILDQELTTFGFPSPSLYEESKRE	AKELNIVAVLNCMKNELLVLRK	NLLAQENV	116				
Qy	121	ETQNLKLGSDMDHLQSCYAKLKEQLETS	SRREMIGLQERDRLOQCN	RSLLHLLKNKDEV	180				
Db	117	ETQNLKPGSDMDHLQSCYAKLKEQLETS	SRREMIGLQERDRLOQCN	RNLHLLKNKDEV	176				
Qy	181	QKLNIIASRATQYNHDKRKERYNKLKER	LHOLVMNKKDKNIAMD	VNLYVGRADKRG	240				
Db	177	QKLNIIASRATQYNHDKRKERYNKLKER	LHOLVMNKKDKNIAMD	VNLYVGRADKRG	236				
Qy	241	SWRTDKTEARNEDMYKILLNDVEYRQKIL	MENAEKLVLOOMKEMISL	SPQKKPR	300				
Db	237	SWRTDKTEARNEDMYKILLNDVEYRQKIL	MENAEKLVLOOMKEMISL	SPQKKPR	296				
Qy	301	ERAEPDGTGTV AISDIEDDSGELSRDS	VWGLSCDTVREQLTNSIRK	QWRILKSHVEKLDNQ	360				
Db	297	ERVDDSTGTV -ISDVEEDAGELSRSD	WDLSCETVREQLTNSIRK	QWRILKSHVEKLDNQ	355				
Qy	361	ASKVHSEGLNEEDVISRQDHEQTEKLE	IEIRCKEMIKAAQQLLOQ	LATTCDTDTSL	420				

THIS PAGE LEFT BLANK

Db 356 VSKVHLGFGNDEVISRDHQETKELELEIOCKEMIKTQOQLLOQLATAYDDTTSL 415  
Qy 421 LRDYLLLEERLKEEWTLPEKOKNFERRRSFTAAIRLGL 464  
Db 416 LRDYLLLEERLKEEWTLPEKOKNFERRRSFTAAIRLGL 459

RESULT 2  
US-09-513-999C-7081  
; Sequence 7081, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7081  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-513-999C-7081

Query Match 19.4%; Score 612.5; DB 2; Length 137;  
Best Local Similarity 87.0%; Pred. No. 1.8e-39;  
Matches 120; Conservative 10; Mismatches 7; Indels 1; Gaps 1;

Qy 272 MENAELKVVLOQMKEMISLLSPKKPRRAEDGTGTVAISDIEDSGELSDSVWGLS 331  
Db 1 MENAELKVVLOQMKEMISLLSPKKPRRAEDGTGTVAISDIEDSGELSDSVWGLS 59  
Qy 332 CDTVREQLTNSIRKQWILKSHVEKLNQSKVHSEGLNEEDVLSRODHQETKELEI 391  
Db 60 CETVREQLTNSIRKQWILKSHVEKLNQSKVHSEGLNEEDVLSRODHQETKELEI 119  
Qy 392 ERCKEMIKAOQLLQOQL 409  
Db 120 QOCKEMIKTQOQLLQOQL 137

RESULT 3  
US-09-949-016-6507  
; Sequence 6507, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6507  
; LENGTH: 3259  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6507

Query Match 7.1%; Score 223.5; DB 2; Length 3259;  
Best Local Similarity 19.9%; Pred. No. 1e-07;  
Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;  
Qy 15 ENKNLSQVTSSTKMS-----PSSLYSQOVLCSVP-LSKNVGHVGVFTCGENIEQ 64  
Db 1225 ENENIGDQLRQIQVRESIDGKLPT--DQESCSSTPGLEELPKATEQHHTQVLES 1282  
Qy 65 SI-----SYLDOELTFGPPSLYE---ESKSEAKRELNIIVAVLNMHELLVLRKNLLA 116  
Db 1283 NLCPDWPSHSEDASALOGGTSVAQIKAKLEAEKVELELKVSTTSELTKKSEVFO 1342  
Qy 117 QESVETONLKLGS--DMDHLOSCYAK-LKEOLETSRRMIGIOERDROLOCKNRSILHQL 173  
Db 1343 QEQINKQGLETESLKTVSHEAEVHAESLOOKLESSQLQIAGL-EHLRELQPK----- 1393  
Qy 174 KNEKDEVOKLQNIIASRATQVNHDKRKERYNKLERLHQLVMNKKDKNIAMVNLVYG 233  
Db 1394 ---LDELQKL-----ISKKEEDVSLSGQL-----SEKEAALTKI----- 1425  
Qy 234 RADGRGSRWRTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK----- 278  
Db 1426 -----QTEIIEQEDLIKALHTOLEMOAKEHDERIKOLOVELCEMKQKPEIGR 1473  
Qy 279 --KVLQOMKEMISLLSPOK-----KKPRRAEDGTGTV-----AISDIED--DSGELS 323  
Db 1474 ESRKQOQIRKLOQAALISRKALKENKSLQBELSLARGTIERLTSLADVESQVSAQNK 1533  
Qy 324 RDSVMG-----LSCDTVR-----EQ 339  
Db 1534 KTVLGRLLALLOERDKLITENDRSILENQSLSSCSLKLALGLETKELVKIESL 1593  
Qy 340 TWS-----IRKQWILKSHVEKLNQSKVHS---EGLNEE-----DVISR 377  
Db 1594 KSKIAESTEWQKHELOKEYEILLQSYENVSNEAERIQHVVEAVROEKQELYGKLRT 1653  
Qy 378 QDHEQETEK---LELEIERCKEMIK-----AQOQLLOQ-----LATTTC 413  
Db 1654 EANKKETEKOLQEAEOEMEMKEMKPKAKSKQKILELEENDRLEAEVHPAGDTAKEC 1713  
Qy 414 DDDTTSLLRDCVLLBEKERLKEEWTLPEKOKNFERRRSFTAAIRLGLERKAFEBERA 473  
Db 1714 ME---TLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSEEVQDL-----KHQIEDNV 1767  
Qy 474 SVWQOFLNMTNPDHONSENVK---LPSAFSGSSDDNLIVHSRPRQKLSHVSANGVPAC 530  
Db 1768 S--KQANLEATE-KHDNQTNTVEGTQSTPGTETEQDSLMSNSTR-----PTC 1811  
Qy 531 TSKLTKSLPASPTS-DFRQTHSCVSEHSSIS-----VLNITPEESKPSSEVAR 577  
Db 1812 SESVPSAKSANPAVSKDF-----SSHDEINNYLQOIQIDOLKERIAGLEBEKQNKPSQ 1864  
Qy 578 ESTDQKWSVQSRPSSREG 595  
Db 1865 TLENERKTLISQLSTKDG 1882

RESULT 4  
US-09-248-796A-16333  
; Sequence 16333, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208

THIS PAGE LEFT BLANK



THIS PAGE LEFT BLANK

87

224 ATGGGAGATTGGATGCTGTTTACAGATCCA-----GAAAGCAAAATATCTCT 271  
Db  
21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Qy  
272 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATCTCAGCAAGTGTATGT 331  
Db  
41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Qy  
332 TCTTCAATACCTTATCGAANAATGTGCACAGTCTTTTTCAGTCGCTTCTGCACAGAAGAT 391  
Db  
61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80  
Qy  
392 AATATTGAACAGAGTATCTCATATCTTGTATCAGGAATTCAGTCTTTTGGTTTCTCTCA 451  
Db  
81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Qy  
452 TTATATGAAGAAATCCAAAGGTAAAGAGCAAGAGAGAGTTAATATATAGTACTGTACTA 511  
Db  
101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120  
Qy  
512 AATTGTATGATGAGTCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAANAATGTG 571  
Db  
121 GluThrGlnAsnLeuLysGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Qy  
572 GAGACACAGAATTTGAAGCCGGAAGTATGATGACCATCTACAGAGCTGCTACTCAAAA 631  
Db  
141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Qy  
632 CTTAAGGAACAACTGGAACCTCCAGGAGGGAATGATGGCTTCAGGAAGAGACAGA 691  
Db  
161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
Qy  
692 CAGTTACAAATGTAAGACAGGAATTTGCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 751  
Db  
181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
Qy  
752 CAAAATTTACAAAATATCATTTGCAAGTCGAGCTACTCAGTATTAATCATGATGAAGAGA 811  
Db  
201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Qy  
812 AAGAGCGGTGAATATAAATACTGAAGGACCTCTACATCACTTGTATGACAGAAA 871  
Db  
221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Qy  
872 GATAGAAAATAGCTATGACATTTTGAATATTATGTCGAGAGCTGATGTAAGAAAGAGGC 931  
Db  
241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Qy  
932 TCCTGGAGGACTGGTAAACCTGAAGCCAGGAATGAAGATGAATGTATAAATTTCTCTTG 991  
Db  
261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280  
Qy  
992 AATGATTTAATAATATCGTCGAACAAATCTTAATGGAATATGCAAGACTTAAGAGGTT 1051  
Db  
281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
Qy  
1052 CTTCAACAAATGAAAAGAAATGATTTCTTCTTCTTCTCCCAAAAGAAAGAAACCTAGA 1111  
Db  
301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
Qy  
1112 GAAAGAGTAGTAGTAGTACAGGAACCTGTT---ATTTCGATGTTGAAAGAAATGCCGGG 1169  
Db  
321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
Qy  
1169 GAACATAGCAGAGAGTAGTATGGGACCTTTCTGTGAACCTGTGAGAGCAGCTTACA 1228  
Db  
341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
Qy  
1229 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGATCATGTAGAAAAGCTTGAATACCAA 1288  
Db  
361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
Qy  
1289 GTTTCAAAGGTACACCTGGAAAGGTTTAAATGATGAGATGTAATCTCACGCAAGACCAT 1348  
Db

381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
Qy  
1349 GAACAGAAACTGAAAAACTCGAGTTAGAAATTCAGCAGCTGTAAAGAAATGATTAAACT 1408  
Db  
401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
Qy  
1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTCGCATATGATGATGATACCACTTCACTA 1468  
Db  
421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluLeuTrpThrLeuPhe 440  
Qy  
1469 TTACGAGAGCTGTTATTTGTTGGAAGAAAAGGAACGCTCTCAAGAGAATGGTCCCTTTT 1528  
Db  
441 LysGluGlnLysAsnPheGluArgGluArgSerPheThrThrGluAlaAlaIleArg 460  
Qy  
1529 AAAGAGCAGAAAAGAAATTTTGAGAGGAGAGAGCAAGCTTTTACAGAAGCCGCTATTTCGC 1588  
Db  
461 LeuGlyLeuGlu----- 464  
Qy  
1589 CTGGGATTTGAGATTTGGCTTTCTCTGCCAGTGTTTTACAGAATAATCTCTTTGGGT 1648  
Db  
464 ----- 464  
Qy  
1649 ATCCATCATTCATCCACGGTCAGTAGTGTGTGTCAGAAGACAGTGCACCAAGACTAT 1708  
Db  
465 -----ArgLysAlaPheGluGluArgAlaSerTrpValLysGlnG 479  
Qy  
1709 ATTTGCTGTTTAAAGAAAGGCAATTTGAAGAGAAAGAGCCAGCTTGGTTAAAGCAGCA 1768  
Db  
479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499  
Qy  
1769 GTTTCTAAATATGACTACTCTTTCACCAACAGACTCAGAAAATGTGAAACTTTTCAGTGC 1828  
Db  
499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLy 519  
Qy  
1829 CTTCTCAGGAAGTCTGATTGGGACAATCTTATAGTGCACTCGAGGCGAGCGCAAAAGAA 1888  
Db  
519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539  
Qy  
1889 GCCTCACAGTGTCTAATGGGTCTCCAGTTTCATGCTCTAAACTTAACTTAACTCTTCC 1948  
Db  
539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559  
Qy  
1949 TGCTTCACTTCCACTTCAGACTTTTCCAGACAGCTCTCTGCATATCTGACATATCTTC 2008  
Db  
559 rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579  
Qy  
2009 AATCAATGTAAGTAAATACTGCTGAAGAAATTAACCAAAATCAGGTGTGGAGGAGAACG 2068  
Db  
579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerG 599  
Qy  
2069 TACAATCAAAAATGGAGTGTGGCTCAGACCTGGATCAGGAAGGTTCTATAGTGG 2128  
Db  
599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615  
Qy  
2129 ATGCTCTGTGAGCTACAAAATTTCTCATGTAGAAAAGATGACTTACCT 2177  
Db

## RESULT 2

US-09-513-999C-3004  
; Sequence 3004, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59 US2 REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681

**THIS PAGE LEFT BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: June 13, 2006, 20:20:55 ; Search time 1459 Seconds  
(without alignments)  
12864.491 Million cell updates/sec  
Title: US-10-644-084-1  
Perfect score: 2692  
Sequence: 1 cgtaggagagtgacaggagc.....ttgattattggagtaatgggg 2692  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 5244920 seqs, 3486124231 residues  
Total number of hits satisfying chosen parameters: 10489840  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_8.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*  
15: Geneseqn2006s.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Query		Match		Length		DB		ID		Description	
Result No.	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID	Score	Match	Length	DB	ID	Description
1	2692	100.0	2692	12	ADO26490	100.0	2692	12	ADO26490	ADO26490	100.0	2692	12	ADO26490	ADO26490	Murine af
2	2189.4	81.3	3195	12	ADO26492	81.3	3195	12	ADO26492	ADO26492	81.3	3195	12	ADO26492	ADO26492	Rat afadi
3	1624.4	60.3	5476	10	ACD19444	60.3	5476	10	ACD19444	ACD19444	60.3	5476	10	ACD19444	ACD19444	cDNA enco
4	1260.4	46.8	2716	4	AAB14625	46.8	2716	4	AAB14625	AAB14625	46.8	2716	4	AAB14625	AAB14625	Human cDN
5	1258.8	46.8	2272	10	ADB63343	46.8	2272	10	ADB63343	ADB63343	46.8	2272	10	ADB63343	ADB63343	Human cDN
6	560	20.8	826	4	AAB06926	20.8	826	4	AAB06926	AAB06926	20.8	826	4	AAB06926	AAB06926	Human cDN
7	441.6	16.4	668	13	ADQ78551	16.4	668	13	ADQ78551	ADQ78551	16.4	668	13	ADQ78551	ADQ78551	Novel can
8	399.8	14.9	6619	5	ABA20636	14.9	6619	5	ABA20636	ABA20636	14.9	6619	5	ABA20636	ABA20636	Human ner
9	399.8	14.9	6619	5	ABA20638	14.9	6619	5	ABA20638	ABA20638	14.9	6619	5	ABA20638	ABA20638	Human ner
10	399.8	14.9	6619	5	ABA20637	14.9	6619	5	ABA20637	ABA20637	14.9	6619	5	ABA20637	ABA20637	Human ner
11	338.8	12.6	1503	4	AAB16996	12.6	1503	4	AAB16996	AAB16996	12.6	1503	4	AAB16996	AAB16996	Human cDN
12	334.2	12.4	563	4	AAH08911	12.4	563	4	AAH08911	AAH08911	12.4	563	4	AAH08911	AAH08911	Human cDN
13	323	12.0	498	2	AAV86670	12.0	498	2	AAV86670	AAV86670	12.0	498	2	AAV86670	AAV86670	EST clone
14	305.8	11.4	403	5	AAF66702	11.4	403	5	AAF66702	AAF66702	11.4	403	5	AAF66702	AAF66702	Novel hum
15	297.2	11.0	428	3	AAAC03006	11.0	428	3	AAAC03006	AAAC03006	11.0	428	3	AAAC03006	AAAC03006	Human sec
16	212	7.9	884	4	AAH06470	7.9	884	4	AAH06470	AAH06470	7.9	884	4	AAH06470	AAH06470	Human cDN
17	212	7.9	2659	4	AAH17820	7.9	2659	4	AAH17820	AAH17820	7.9	2659	4	AAH17820	AAH17820	Human cDN
18	211.4	7.9	833	4	AAH03373	7.9	833	4	AAH03373	AAH03373	7.9	833	4	AAH03373	AAH03373	Human cDN

c 19	130.6	4.9	513	12	ACH72399	Ach72399	Human gen
c 20	123.6	4.6	178	12	ACH86155	Ach86155	Human gen
c 21	111.4	4.1	441	6	ABN18923	ABN18923	Human ORF
c 22	64.8	2.4	180385	10	ADL13931	ADL13931	Osteoarthe
c 23	61.4	2.3	2000	8	ADA71938	ADA71938	Rice gene
c 24	61.2	2.3	30191	12	ADQ97654	ADQ97654	Mouse can
c 25	60.4	2.2	621	5	AAS90688	AAS90688	DNA enco
c 26	59.2	2.2	110000	14	ADZ12814_2	ADZ12814_2	Continuation (3 of
c 27	59	2.2	2395	15	AEE62999	AEE62999	Leucocyto
c 28	58.2	2.2	198	3	AAC25026	AAC25026	Human sec
c 29	57.6	2.1	3399	2	AAT05868	AAT05868	Chicken 1
c 30	57.4	2.1	1556	14	AEB67285	AEB67285	Rice geno
c 31	56.8	2.1	15507	8	AAL53680	AAL53680	Genomic D
c 32	56.4	2.1	1686	2	AAQ87587	AAQ87587	DNA enco
c 33	55.8	2.1	2093	10	ADB63512	ADB63512	Human cDN
c 34	54.6	2.0	13407	13	ABD33029	ABD33029	Mouse can
c 35	54	2.0	1199	13	ADX27288	ADX27288	Plant ful
c 36	52.4	1.9	1062	13	ADR64454	ADR64454	Cotton CD
c 37	52.2	1.9	1149	3	AAC46459	AAC46459	Arabidops
c 38	50.8	1.9	556	12	ADP91525	ADP91525	Cotton ex
c 39	50.8	1.9	193853	11	ACN44956	ACN44956	Mouse gen
c 40	50.6	1.9	1997	13	ADR07778	ADR07778	Full leng
c 41	50.6	1.9	235962	14	ADZ12926	ADZ12926	Murine ca
c 42	50	1.9	930	13	ADU02094	ADU02094	Novel hum
c 43	49.8	1.8	12025	6	ABL33298	ABL33298	Human imm
c 44	49.8	1.8	19380	6	AAS61426	AAS61426	Human gen
c 45	49.4	1.8	1761	13	ADR07771	ADR07771	Full leng

ALIGNMENTS

RESULT 1  
ADO26490  
ID ADO26490 standard; DNA; 2692 BP.  
AC ADO26490;  
XX  
DT 29-JUL-2004 (first entry)  
DE Murine afadin dilution domain binding protein (ADIP) DNA SeqID 1.  
XX  
KW murine; mouse; gene; ds; afadin dilution domain binding protein; ADIP;  
KW afadin; actinin; binding inhibitor; cardiant; heat disease;  
KW myocardial infarction; myocarditis.  
XX  
OS Mus musculus.  
FH Key Location/Qualifiers  
FT CDS 80..1927  
FT /\*tag= a  
FT /\*product= "ADIP protein"  
XX  
PN JP2004135658-A.  
XX  
PD 13-MAY-2004.  
XX  
PF 14-AUG-2003; 2003JP-00293554.  
XX  
PR 27-SEP-2002; 2002JP-00284263.  
XX  
PA (EISA ) EISAI CO LTD.  
XX  
DR WPI; 2004-404616/38.  
DR P-PSDB; ADO26491.  
XX  
PT New polynucleotide encoding an afadin dilution domain binding protein  
PT having avidity with afadin or actinin, useful for diagnosing heart  
XX diseases e.g. myocardial infarction.  
PS Claim 1; SEQ ID NO 1; 37pp; Japanese.  
CC This invention relates to a novel isolated nucleic acid encoding an

CC	afadin dilution domain binding protein (ADIP) that exhibits an avidity														
CC	with afadin/actinin. Specifically, it refers to screening assays to														
CC	identify compounds that modulate ADIP avidity and provides suitable														
CC	agonists, antagonists and antibodies thereof. The present invention														
CC	provides methods to identify afadin and actinin binding inhibitors														
CC	therapeutically as cardiants to diagnose and/ or treat heat disease such														
CC	as myocardial infarction or myocarditis. This polynucleotide sequence is														
CC	the murine ADIP DNA sequence of the invention.														
XX															
SQ	Sequence 2692 BP; 814 A; 525 C; 691 G; 662 T; 0 U; 0 Other;														
	Query Match	100.0%;	Score 2692;	DB 12;	Length 2692;										
	Best Local Similarity	100.0%;	Pred. No. 0;												
	Matches 2692;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;										
Qy	1	CGTAGGAGGTGACAGGAGCTGTTGTAAGCGTTCGACAGCTGAGCGCGCTCCTCAGGTAT	60												
Db	1	CGTAGGAGGTGACAGGAGCTGTTGTAAGCGTTCGACAGCTGAGCGCGCTCCTCAGGTAT	60												
Qy	61	CCTGGCTCTGGAACTTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTATC	120												
Db	61	CCTGGCTCTGGAACTTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTATC	120												
Qy	121	AGAAAAAATAATCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTGA	180												
Db	121	AGAAAAAATAATCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTGA	180												
Qy	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGTTTTCGG	240												
Db	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGTTTTCGG	240												
Qy	241	TGCTCTTCGACAGGAGAGAACATTTGAACAAAGTATTTCCATCTTGATCAGGAGCTGAC	300												
Db	241	TGCTCTTCGACAGGAGAGAACATTTGAACAAAGTATTTCCATCTTGATCAGGAGCTGAC	300												
Qy	301	CACCTTCGGGTTTCTTGTATGAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT	360												
Db	301	CACCTTCGGGTTTCTTGTATGAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT	360												
Qy	361	AAATATAGTCGTGTTCTGTAATGAAAGAGCTGCTCGTCTTCAGCGGAGAAACCT	420												
Db	361	AAATATAGTCGTGTTCTGTAATGAAAGAGCTGCTCGTCTTCAGCGGAGAAACCT	420												
Qy	421	GCTGGCCAGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGACATGGACCACT	480												
Db	421	GCTGGCCAGAGAGCGTGGAGACACAGAACTTGAAGCTGGGAGTGACATGGACCACT	480												
Qy	481	GCAGAGCTGTACGCCAAAATTAAAGGAGCAGTTGGAAAACGTCACGCGGAGATGATCGG	540												
Db	481	GCAGAGCTGTACGCCAAAATTAAAGGAGCAGTTGGAAAACGTCACGCGGAGATGATCGG	540												
Qy	541	GCTTCAAGAGAGACAGGAGCTGCAGTCCAGAAACAGGAGTTTGATCAGCTCCTGAA	600												
Db	541	GCTTCAAGAGAGACAGGAGCTGCAGTCCAGAAACAGGAGTTTGATCAGCTCCTGAA	600												
Qy	601	GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA	660												
Db	601	GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA	660												
Qy	661	TAATCATGATGTGAAGAGGAGGCGTGAATATAAAGCTAAAGGAGCGCTGCATCA	720												
Db	661	TAATCATGATGTGAAGAGGAGGCGTGAATATAAAGCTAAAGGAGCGCTGCATCA	720												
Qy	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTAAATATGTGGGTG	780												
Db	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTAAATATGTGGGTG	780												
Qy	781	AGCTGATGGCAACGAGGCTCATGGAGACTGACAAAACAGAGCAGGAGATGAAGATGA	840												
Db	781	AGCTGATGGCAACGAGGCTCATGGAGACTGACAAAACAGAGCAGGAGATGAAGATGA	840												
Qy	841	GATGTACAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAGCAGATCTCTGATGGAG	900												

841	GATGTACAAAATTTCTGTTGAATGATTATGATAGTACCGCCAGAGCAGATCTCTGATGGAGAA	900
901	CGGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	960
901	CGGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	960
961	TCAGAAGAAGAGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1020
961	TCAGAAGAAGAGCCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1020
1021	TATAGAAGATGACTCTGGGAACTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1080
1021	TATAGAAGATGACTCTGGGAACTAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1080
1081	TGTGAGAGACAGCTGACAAACAGCATCAGAAACAGTGGAGATTTTGAAGAAGCTCATGT	1140
1081	TGTGAGAGACAGCTGACAAACAGCATCAGAAACAGTGGAGATTTTGAAGAAGCTCATGT	1140
1141	AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT	1200
1141	AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT	1200
1201	CATCTCAGCAAGACCATGAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTG	1260
1201	CATCTCAGCAAGACCATGAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTG	1260
1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGCCACCACTGTGA	1320
1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGCCACCACTGTGA	1320
1321	TGATGACACCACTCACTCTGAGAGACTGTTACTCTGCTGGAAGAAAGGAAACCCCTTAA	1380
1321	TGATGACACCACTCACTCTGAGAGACTGTTACTCTGCTGGAAGAAAGGAAACCCCTTAA	1380
1381	AGAAGAGTGAGCCCTTTTAAAGAGCAAAAAAGATTTTGAAGAGAGAGAGAGAGAGCTT	1440
1381	AGAAGAGTGAGCCCTTTTAAAGAGCAAAAAAGATTTTGAAGAGAGAGAGAGAGAGCTT	1440
1441	TACAGAGCTGCCATTCGATTGGGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1500
1441	TACAGAGCTGCCATTCGATTGGGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1500
1501	CTGGTAAAGCAGCAGCTTTTAAACATGACAACTTTTGACCAACAGAGAACTCAGAAAAATGT	1560
1501	CTGGTAAAGCAGCAGCTTTTAAACATGACAACTTTTGACCAACAGAGAACTCAGAAAAATGT	1560
1561	GAAACTTTTCAGTGCCCTTCTCAGGAAGTTCTGATCCAGACAACTTTATAGTCCACTCAG	1620
1561	GAAACTTTTCAGTGCCCTTCTCAGGAAGTTCTGATCCAGACAACTTTATAGTCCACTCAG	1620
1621	GCCACGCAAG	1680
1621	GCCACGCAAG	1680
1681	GACTTAAATCTCTCTGCTCCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1740
1681	GACTTAAATCTCTCTGCTCCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1740
1741	GTCGTAACACAGTTTCCATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1800
1741	GTCGTAACACAGTTTCCATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1800
1801	GGTTGCAAG	1860
1801	GGTTGCAAG	1860
1861	GGGGTGTCTACAGGGAGTCTCTCTCGGCTTTCAGGAGCGCTCAAGGGAGCCGAGATGACTT	1920
1861	GGGGTGTCTACAGGGAGTCTCTCTCGGCTTTCAGGAGCGCTCAAGGGAGCCGAGATGACTT	1920
1921	ACCTTAAATGTGGGGCTCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1980
1921	ACCTTAAATGTGGGGCTCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1980

```
QY 1981 GTGTAGCATAAAGTCAGTCTTAAGTAAAGTCTCAGAGTGTGTTGTTGGACTTCGC 2040
Db |||||
QY 1981 GTGTAGCATAAAGTCAGTCTTAAGTAAAGTCTCAGAGTGTGTTGTTGGACTTCGC 2040
Db |||||
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAGAGTGCAGAGCTTTGGTTGT 2100
Db |||||
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAGAGTGCAGAGCTTTGGTTGT 2100
Db |||||
QY 2101 GTGTAGTAAAGAGCCCTGGCTCTGTGCTGCTGAGTGCAGGAATGCATGGCGTTTGGATGAA 2160
Db |||||
QY 2101 GTGTAGTAAAGAGCCCTGGCTCTGTGCTGCTGAGTGCAGGAATGCATGGCGTTTGGATGAA 2160
Db |||||
QY 2161 ACAGAGCGCTGGAATGATGCTCGCCAGGTACCGAGAGAGCACTTTAGGAGCTGGT 2220
Db |||||
QY 2161 ACAGAGCGCTGGAATGATGCTCGCCAGGTACCGAGAGAGCACTTTAGGAGCTGGT 2220
Db |||||
QY 2221 TCTCTAAACATTAATATTCCTCCCAAGTGTGGTGGCAATGGAAGTGTAGCCTTTACT 2280
Db |||||
QY 2221 TCTCTAAACATTAATATTCCTCCCAAGTGTGGTGGCAATGGAAGTGTAGCCTTTACT 2280
Db |||||
QY 2281 TGAATGTATACGTAGATTTTAAACAAAGCAGGTCTATATTTATTTAGTGTGAT 2340
Db |||||
QY 2281 TGAATGTATACGTAGATTTTAAACAAAGCAGGTCTATATTTATTTAGTGTGAT 2340
Db |||||
QY 2341 TTTGGGATACCTCTTTCATATGTTTGTGCTGTACATAATATACATGACTATGTTAA 2400
Db |||||
QY 2341 TTTGGGATACCTCTTTCATATGTTTGTGCTGTACATAATATACATGACTATGTTAA 2400
Db |||||
QY 2401 GAGGCTTTAAGGTTTAAACCTTCACACCATGCTTGGATATAGCATTTTCATGCCAATTA 2460
Db |||||
QY 2401 GAGGCTTTAAGGTTTAAACCTTCACACCATGCTTGGATATAGCATTTTCATGCCAATTA 2460
Db |||||
QY 2461 AATGTTTTCAGTGGCATGGTGTACAGAGGTTAGGACCACTGCCACATGACAGTTAAGA 2520
Db |||||
QY 2521 CTTTATTTTAAAGCATCTGGCAATTAATTTCAAGCCCTTTCATAGCTGAGTTTCAAG 2580
Db |||||
QY 2521 CTTTATTTTAAAGCATCTGGCAATTAATTTCAAGCCCTTTCATAGCTGAGTTTCAAG 2580
Db |||||
QY 2581 ATAAGTAACTACTAAAGCTTACATTTTGTGATTTTAAAGCATTTTATTTTATTTAT 2640
Db |||||
QY 2581 ATAAGTAACTACTAAAGCTTACATTTTGTGATTTTAAAGCATTTTATTTTATTTAT 2640
Db |||||
QY 2641 ATATGTAATGTTATTAATTTCTAAGAGGAATATGATTATGGAGTAATGGGG 2692
Db |||||
QY 2641 ATATGTAATGTTATTAATTTCTAAGAGGAATATGATTATGGAGTAATGGGG 2692
Db |||||
```

## RESULT 2

ADO26492  
ID ADO26492 standard; DNA; 3195 BP.

XX AC ADO26492;

XX XX

XX DT 29-JUL-2004 (first entry)

XX XX

XX DE Rat afadin dilution domain binding protein (ADIP) DNA SeqID 3.

XX XX

XX KW rat; gene; ds; afadin dilution domain binding protein; ADIP; afadin;  
KW actinin; binding inhibitor; cardiant; heat disease;

XX KW myocardial infarction; myocarditis.

XX XX

XX OS Rattus norvegicus.

XX XX

XX FH Key Location/Qualifiers

XX FT CDS 79..1920

XX FT /\*tag= a

XX FT /product= "ADIP protein"

XX XX

XX PN JP2004135658-A.

XX XX

PD 13-MAY-2004.

XX XX

PF 14-AUG-2003; 2003JP-00293554.

XX XX

PR 27-SEP-2002; 2002JP-00284263.

XX XX

PA (EISA ) EISAI CO LTD.

XX XX

DR WPI; 2004-404616/38.

XX XX

DR P-PSDB; ADO26493.

XX XX

PT New polynucleotide encoding an afadin dilution domain binding protein  
having avidity with afadin or actinin, useful for diagnosing heart

PT diseases e.g. myocardial infarction.

XX XX

PS Claim 1; SEQ ID NO 3; 37pp; Japanese.

XX XX

CC This invention relates to a novel isolated nucleic acid encoding an  
afadin dilution domain binding protein (ADIP) that exhibits an avidity  
with afadin/actinin. Specifically, it refers to screening assays to  
identify compounds that modulate ADIP avidity and provides suitable  
agonists, antagonists and antibodies thereof. The present invention  
provides methods to identify afadin and actinin binding inhibitors  
therapeutically as cardiants to diagnose and/ or treat heart disease such  
as myocardial infarction or myocarditis. This polynucleotide sequence is  
the rat ADIP DNA sequence of the invention.

XX XX

SQ Sequence 3195 BP; 929 A; 648 C; 794 G; 823 T; 0 U; 1 Other;

Query Match 81.3%; Score 2189.4; DB 12; Length 3195;

Best Local Similarity 90.1%; Pred. No. 0;

Matches 2427; Conservative 0; Mismatches 247; Indels 20; Gaps 7;

QY 2 GTAGGAGAGTGCACAGGAGCTGTGTAAGCGTCGACACTGAGCGCTCCTCAGGTATC 61

Db 1 GTAGGAGAGTGCACAGGAGCTGTGTAAGCGTCGACACTGAGCGCTCCTCAGGTATC 60

QY 62 CTGCTCTGGAACTTGTCTATGGGAGATTTGGATGACTGTGACAGATCCAGTTCTGTGTACA 121

Db 61 CTGCTCTGGAACTTGTCTATGGGAGATTTGGATGACTGTGACAGATCCAGTTCTGTGTACA 120

QY 122 GAAACAAAATCTCTCTCAATATACCTCAGAAAACAAAGATGTCTCCGTCAGTTTGTAC 181

Db 121 GAAACAAAATCTCTCTCAATATACCTCAGAAAACAAAGATGTCTCCGTCAGTTTGTAC 180

QY 182 TCCAGCAAGTCTGTGCTCTTCACTTCTAGTACCTTTATCCAAAACAGTCATGTTTTCGGT 241

Db 181 TCCAGCAAGTCTGTGCTCTTCACTTCTAGTACCTTTATCCAAAACAGTCATGTTTTCAGT 240

QY 242 GTCTTCTGCACAGGAGAGAACATTGAAACAAAGTATTTCTATCTTGTATCAGGAGCTGACC 301

Db 241 GCCTTCTGCACAGGAGAGAACATTGAAACAAAGTATTTCTATCTTGTATCAGGAGCTGACC 300

QY 302 ACCTTCGGGTTTCCCTTCTGTAATGAGATCCAAAAGTAAAGGCGGAGAGAGATTA 361

Db 301 ACCTTCGGGTTTCCCTTCTGTAATGAGATCCAAAAGTAAAGGCGGAGAGAGATTA 360

QY 362 AATATAGTCTGCTGTTCTGAACTGTATGAAACAGCTGTCTGCTTTCAGCGGAAAGACCTG 421

Db 361 AATATAGTCTGCTGTTCTGAACTGTATGAAACAGCTGTCTGCTTTCAGCGGAAAGACCTG 420

QY 422 CTGCCCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGCATCGACACCTG 481

Db 421 CTGCCCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGCATCGACACCTG 480

QY 482 CAGAGCTGCTACGCCCAACTTAAAGGAGCTTGAAGAGCTTCCAGCGGAGAGATATCGGG 541

Db 481 CAGAGCTGCTACGCCCAACTTAAAGGAGCTTGAAGAGCTTCCAGCGGAGAGATATCGAG 540

QY 542 CTTCAAGAGAGAGACAGGAGCTGACGTGCAAGAACAGGAGTTTGTACATCAGCTCCTGAAG 601

Db 541 CTTCAAGAGAGAGACAGGAGCTGACGTGCAAGAACAGGAGTTTGTACATCAGCTCCTGAA 600







PA	(HELI-) HELIX RES INST.
XX	Ota T, Ieogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	WPI; 2001-318749/34.
DR	
XX	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT	
PT	
XX	
PS	Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.
XX	
CC	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX	
SQ	Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;
	Query Match            46.8%; Score 1260.4; DB 4; Length 2716;
	Best Local Similarity   79.7%; Pred. No. 0;
	Matches 1716; Conservative   0; Mismatches 376; Indels 60; Gaps 17;
Qy	594 TCCTGGAAGATGAGAAAGATGAGGTACAAAATTACAAAATATCATAGCCCGGGCTA 653
Dd	1 TACTAAAGAATGAGAAAGATGAGGTGC AAAATTTACAAAATATCATTCGAAGTCAGCTA 60
Qy	654 CTCNGTATAATCATGTGTGAAGAGGAGCGTGTAATATAATAGCTAAAGAGGCC 713
Dd	61 CTCAGTATAATCATGTATGAAGAGAAAAGAGCGGTGAATATAATAAATCGAAGAACGTC 120
Qy	714 TGCAATCAGCTCGTTATGACAAGAAGGATAAAAACATAGCATGGATGTTTTAAATTTATG 773
Dd	121 TACATCAACTTGTATTGACAAGAAGAGATAGAAAAATAGCTATGGAATTTTGAATTTATG 180
Qy	774 TGGGTTCGAGCTGTATGGCAAAACGAGGCTCATGGAGGACTGACAAAACAGAGCCAGGAATG 833
Dd	181 TCGGGAGAGCTGATGGA AAAAGAGGCTCTCTGGAGACTGGTAAACTGAAGCCAGGAATG 240
Qy	834 AGATGAGATGTTACAAAATTTCTGTTGAATGATTATAGTACCGCCAGAAAGAGATCTCTGA 893
Dd	241 AAGATGAAATGTTATAAAATTTCTCTTGAATGATTATGAATATCGTCAGAAACAAATCTCAA 300
Qy	894 TGGAGACGCGGAGCTCAAGAAGGTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCC 953
Dd	301 TTGGAATTCAGAACTTAAAGAGGTTCTTCAACAAATGAAAAGGAATGATTTCTCTTC 360
Qy	954 TGTTCTCTCAGAGAAGAACCCAGGGAAAGAGCAGAGGACGGCACAGGCACTGTTGCTA 1013
Dd	361 TTTCTCCCCAAGAAGAAACCTAGAGAGAGTAGATGATAGTACAGGAACGT---TA 417
Qy	1014 TTCCTCGATATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGGCTTTCTCT 1073







Qy	421	GCTGCCCCAGGAGACGCTGGAGACACAGAACTTGNAGCTGGCAGTGCATGGACCACCT	480
Db	553	TCTAGCTCAGGAAATGTGGAGACACAGAA'TTGAAGCCGGGAAGTGTATATGGAACCATCT	612
Qy	481	CGAGAGCTGCTACGCGCAAAC'TTAAGAGGACAGTGTGGAAACGTC'CCAGGCGGGAGATGATCGG	540
Db	613	ACAGAGCTGCTACTCAAACTTAAAGRACAACTGGAAACCTCCAGGAGGGAATGATTGG	672
Qy	541	GCTTCAGAGAGACAGGACAGCTGCAGTGCAGAAACAGGAGT'TTGCATCAGCTCCTGAA	600
Db	673	GCTTCAGGAAAGAGACAGACAGT'TACAATGTGAAGAACAGGAAT'TTGCATCAGCTACTAAA	732
Qy	601	GAATGAGAAAGATGAGGTAFCAAAAATTA'CAAAAATATCATGCCAGCGGGCTACTCAGTA	660
Db	733	GAATGAGAAAGATGAGGTGCAAAAATTA'CAAAAATATCAIT'GCAAGTCGAGCTACTCAGTA	792
Qy	661	TAATCATGATGTGAAGAGGAAGGAGCGTGAATATA'ATAAGCTAAAGGAGCGCCTGCATCA	720
Db	793	TAATCATGATATGAAGAGAAAGAGCGTGAATATA'ATAA'CTGAAGGACGCTCTACATCA	852
Qy	721	GCTCGTTATGAACAAGAAGGATATAAACATAGCOATGGATGTTT'TAAATTAATGTGGGTG	780
Db	853	ACTTGT'TATGAACAAGAAAGATAAGAAAATAGCTATG'GACATTTTGAATTAATGTCCGGAG	912
Qy	781	AGCTGATGGCAAAACGAGCTCATCGAGGACTGACAAACAGAACCCAGGAATGAAGATGA	840
Db	913	AGCTGATGGAAAAGAGGCTCCTGGAGACTTGGTAAAC'CTGAGCCAGGAATGAAGATGA	972
Qy	841	GATGTACAAAATTCCTGTTGAATGATTA'TAGTATGCCGACAGACAGAT'CCTGATGGAGAA	900
Db	973	AATGTATAAAATTCCTTGAATGATTA'GATATCGTCAGAAACAAATCCTAATGGAAA	1032
Qy	901	CGCGGAGCTGNAAGAGTCTCTCAGCAGATGAAGAAGAGATGAT'CTCTCTGTCTCC	960
Db	1033	TGCAGAACTTAAAGAGGTTCTTCAACAAATGAAA'AAAGAAATGATTTCTCTCTTCTCTCC	1092
Qy	961	TCAGAGAAAGAGCCCAAGGAAAGACAGAGAGCGCACAGGCACTGT'GTCTATCTCCGA	1020
Db	1093	CCAAAAGAGNAACCTAGAGNAAGGTAGATGATAGTACAGGA'AACTGT---TATTTCCGA	1149
Qy	1021	TATAGAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGGCCCTTTCTGTGACAC	1080
Db	1150	TGTTGAAGAAGATGCCCGGGAACTAAGCAGAGAGAGTATGTGGGACCTTTTCTGTGAAC	1209
Qy	1081	TGTCAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAA'TTTTGAAGTCAATG	1140
Db	1210	TGTGAGAGAGCAGCTTACAAACAGCATCAGAAAACAGTGGAGAA'TTTTGAAGTCAATG	1269
Qy	1141	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGCC'TTAATCAGGAGGACGT	1200
Db	1270	AGAAAAGCTTGATAACCAAGTTTCAAGGTTACACTCTGAAAGT'TTTAATGATGAAGATG	1329
Qy	1201	CATCTCAGCAACAAGCCATGAGCAAGAGACTGAGAA'CTGGAGCTGGAGATTCAGCGGTG	1260
Db	1330	AATCTCAGCAACAAGCCATGAACAAGAAAC'TGAAAAA'CTCGAGTTAGAAATTCAGCAGTG	1389
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGGAGCTGCCACACAGTGTGA	1320
Db	1390	TAAAGAATGATTA'AAA'CTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGTCATATGA	1449
Qy	1321	TGATGACACACACCTCACTGTTTGGCAGACTGTTACTCTGCTGGAAGAAAGCAAGCCCTTAA	1380
Db	1450	TGATGATACCACTTCACTATTACAGACTGTTATTGTTGGAAAGAAAGCAAGCTCTCAA	1509
Qy	1381	AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAA'TTTTGAAGAGAAAGGCGGAGCTTT	1440
Db	1510	AGAAGAATGGTCCCTTTTAAAGAGCAGAAAAAGAA'TTTTGAAGAGGAGAGACGAAGCTT	1569
Qy	1441	TACAGAGCTGCCCATTCGATTGGGTTGGAG-----	1471
Db	1570	TACAGAGCCGCTATTCCGCTGGGAATGGAGATTGGCTTTCTCCTGCCAAGTGTTTTACA	1629

Qy	1472	-----1471	
Db	1630	CAAGAAATCTCTTTTGGGTATCCATCATCCACGGTCAGTAGGTGTGGTGTGCAAG	1689
Qy	1472	-----AGAAAGGCGTTTGAAGAGAGAGCGAG	1496
Db	1690	ACAGTGCACCAAGACTATATTGCTGTTTAAAGAAAGAGCATTTGAAGAGAAAGAG	1749
Qy	1497	CCAGCTGGGTAAAGCAGCAGTTTTTAAACATGACGAATTTTGACCACGAACTCAGAAA	1556
Db	1750	CCAGTGGTTAAAGCAGCAGTTTCTAAATATGACTACCTTTTGACCACGAACTCAGAAA	1809
Qy	1557	ATGTGAACCTTTTCAGTGCCTTCTCAGGAAGTCTCTGATCCAGACAATCTTTATAGTCCACT	1616
Db	1810	ATGTGAACCTTTTCAGTGCCTTCTCAGGAAGTCTCTGATGGGCAATCTTTATAGTGCAC	1869
Qy	1617	CACGGCCACGGCAAAAGACTACACAGTCTGCTTAATGGGTGCCAGCTTGACACATCAA	1676
Db	1870	CGAGCAGCGCAAAAGAGCCTCACAGTGTGCTTAATGGGTTCACAGTTTGCAATGTCTA	1929
Qy	1677	AACTGACTAAATCTCTTCCTCGCCTCACCTTCTACTTCAGACTTTTGGCCAGACACATTCAT	1736
Db	1930	AACTTACTAAATCTCTTCCTGCTTCACTTCCACTTCAGACTTTTGGCCAGACAGTTCT	1989
Qy	1737	GTGTGCTGAACACAGTTCCATCAGTGTGCTGAATATAACTCTCTGAAGAAAGTAAACCAA	1796
Db	1990	GCATATCTGAACATAGTTCAATCAATGACTGAATATAACTGCTGAAGAAATTTAAACCAA	2049
Qy	1797	GTGAGTTGCAAGAGAAGACCGATCAGAACTGGAGCGTGCAGTCGAGGCCAGCTCGC	1856
Db	2050	ATCAGGTTGGAGGAGAACGTACAAATCAAAAATGGAGTGTGGCGTCAAGACCTGGATCAC	2109
Qy	1857	GGAGGGGTGCTACAGCGGATGCTCTCGGCCCTTCAGGAGCGCTCACGGGGACCGAGATG	1916
Db	2110	AGNAGGTTGCTATAGTGGATGCTCTTGAGCTACACAAATTCATGTAAGAAAAGATG	2169
Qy	1917	ACTTACCTTAAATGTGCGGCTGCAGTGTGTTCCAGATGTGCGCTAGAGG-----	1968
Db	2170	ACTTACCTTAGACATGTGGACTGGAATTTTTTTTCAATTAATGTGTTCAATCAAGTTTCACAT	2229
Qy	1969	---AGTTGACACAGGTTGTAGCATAAAGTCAGTCGTCT	2003
Db	2230	CTAAGTTGAACACAGGGTGTGCTCAATAAGTCAGTTATCT	2267
RESULT 6			
AAH06926			
ID	AAH06926 standard; cDNA; 826 BP.		
XX	AAH06926;		
XX	26-JUN-2001 (first entry)		
DT	Human cDNA clone (5'-primer) SEQ ID NO:3761.		
XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.		
KW	Homo sapiens.		
XX	EP1074617-A2.		
PN	07-FEB-2001.		
XX	28-JUL-2000; 2000EP-00116126.		
PF	29-JUL-1999; 99JP-00248036.		
XX	27-AUG-1999; 99JP-00300253.		
PR	11-JAN-2000; 2000JP-00118776.		
PR	02-MAY-2000; 2000JP-00183767.		
PR	09-JUN-2000; 2000JP-00241899.		
XX	(HELI-) HELIX RES INST.		
XX			

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.  
XX  
XX Claim 1; SEQ ID NO 3761; 2537pp + Sequence Listing; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602 full-  
CC length cDNAs defined in the specification. Where a primer set comprises:  
CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
CC complementary strand of a polynucleotide which comprises one of the 5602  
CC nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
XX present invention

XX  
SQ Sequence 826 BP; 329 A; 120 C; 179 G; 195 T; 0 U; 3 Other;

Query Match 20.8%; Score 560; DB 4; Length 826;

Best Local Similarity 82.1%; Pred. No. 3.3e-142;

Matches 680; Conservative 0; Mismatches 142; Indels 6; Gaps 3;

QY 594 TCCTGAAGATGAGAAGATGAGGTACAAAATATACAAAATATCATAGCCAGCGGGCTA 653  
DB 1 TACTAAAGATGAGAAGATGAGGTGCAAAAATATACAAAATATCATGCAAGTCGAGCTA 60  
QY 654 CTCAGTATATCATGTGTAAGAGGAGGAGCGTGAATATATAAGCTAAGAGGCGCC 713  
DB 61 CTCAGTATATCATGTGTAAGAGGAGGAGCGTGAATATATAAGCTAAGAGGAGCGTC 120  
QY 714 TGCATCAGCTGTTATGAACAAGAGGATGAACATATGATCCATGATGTTTAAATATG 773  
DB 121 TACATCAACTGTTATGAACAAGAGGATGAACATATGATCCATGATGTTTAAATATG 180  
QY 774 TGGGTGCGAGCTGATGCGAAACGAGGCTCATGGAGGACTGACAAAACAGAGCCAGGAATG 833  
DB 181 TCGGAGAGCTGATGGAAAGAGGCTCTCTGAGGAGCTGGTAAACTGAAGCCAGGAATG 240  
QY 834 AAGATGAGATGACAAAATCTCTGTTGAATGATATGATACCGCCAGAGAGAGATCTGA 893  
DB 241 AAGATGAAATGTATAAAATCTCTTGAATGATATGATATATCGTCAGAAAACAAATCCCTAA 300  
QY 894 TGGAGACGCGAGCTGAAGAAGGTCTCTCAGCAGATGAAGAAGGAGATGATCTCTCC 953  
DB 301 TGGAAATGCAAGACTTGAAGGTTCTTCAAAATGAAGAAAGAAATGATTTCTCTTC 360  
QY 954 TGTCTCTCTCAGAAGAAGAGCCCGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013  
DB 361 TTTCTCCCAAAAGAGAAACCTAGAGAAAGAGTAGATGATAGTACAGGAAGTGT---TA 417  
QY 1014 TCTCGATATAGAAGATGACTCTGGGGAACCTGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1073  
DB 418 TTTCCGATGTTGAAGAAGATCCGGGGAACCTAAGCAGAGAGAGATATGAGGAGCCTTTCT 477

QY 1074 GTGACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGA 1133  
DB 478 GTGAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGA 537  
QY 1134 GTCATGTAGAAAACTCGATAACCAAGCTTCGAAAGGTACACTCAGAGGCGCTTAAATGAGG 1193  
DB 538 GTCATGTAGAAAAAGCTTGATAACCAAGCTTCAAAGGTACACCTCGAAGGTTTAAATGATG 597  
QY 1194 AGACGCTCATCTCAGCAACAGACGTCGCAAGAGACTGAGAACTCGAGCTGAGATTG 1253  
DB 598 AAGATGTAATCTCAGCAACAGACCATGAACAAGAACTGAAAAAACTCGAGTTAGAAATTC 657  
QY 1254 AGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCCACCA 1313  
DB 658 AGCAGTGTAAAGAAATGATTAACCTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTG 717  
QY 1314 CGTGTGATGATGACACACCTCATCTGTTGGAGACTGTTACTT-GCTGGAGAAAGGAA 1372  
DB 718 CATATGATGATG--ATCCACTTCTATTACGAGACTGTTATTTGGTTGGAGAAAGGAC 775  
QY 1373 CGCCTTAAAGAGAGTGGACCTTTTAAAGACCAAAAAGAAATTT 1420  
DB 776 CGTNTCAAGAAAAAATGGGCCCTTTTAAAGAACCCNAAAAGAAAT 823  
RESULT 7  
ADQ78551  
ID ADQ78551 standard; DNA; 668 BP.  
XX  
AC ADQ78551;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Novel canine microarray-related DNA sequence SeqID10110.  
XX  
KW canine microarray; drug screening; toxicity assay;  
KW environmental pollutant; cellular response; gene expression profile;  
KW toxic response; liver necrosis; fatty liver disease;  
KW protein adduct formation; hepatitis; dog; ds.  
XX  
OS Canis familiaris.  
XX  
PN WO2004063324-A2.  
XX  
PD 29-JUL-2004.  
XX  
PF 05-MAY-2003; 2003WO-US013853.  
XX  
PR 03-MAY-2002; 2002US-0377240P.  
XX  
PA (GENE-) GENE LOGIC INC.  
PA (PFIZ ) PFIZER PROD INC.  
XX  
PI Diggins JC, Porter M, Wei T;  
XX WPI; 2004-561890/54.  
XX  
XX Claim 1; SEQ ID NO 10110; 41pp; English.  
XX  
CC This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and  
CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers

CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver,  
CC heart, brain or testes, or other pathologies associated with at least one  
CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention. Note: The sequence  
CC data for this patent does not form part of the printed specification but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 668 BP; 205 A; 148 C; 141 G; 157 T; 0 U; 17 Other;

Query Match		16.4%;	Score 441.6;	DB 13;	Length 668;
Best Local Similarity		79.6%;	Pred. No. 7.8e-110;		
Matches 510;		Conservative 0;	Mismatches 131;	Indels 0;	Gaps 0;
QY	1310	ACCAGCTGTGATGATGACACACACCTCTACTGTTGCGAGACTGTTACTGTCGGAAGAAAG	1369		
DB	1	ACAGATCGGATGATGACANACGCTCTGTCGCGAGACTGTTACTGCNNNNNAGAA	60		
QY	1370	GAACCCCTTAAGAAGAGATGGACCCCTTTTAAAGAGACAAAGAAATTTTGAGAGAA	1429		
DB	61	NNNNNCTCAANNAGATGGTNCCTGTTTAAAGAGACAAAGAAATTTGAGAGAA	120		
QY	1430	AGGCGAAGCTTTACAGAGCTGCCATTCGATTTGGGTTGGAGAAAGCGGTTTGAAGAA	1489		
DB	121	AGACGAAGCTTTACAGAGAGCGACCATTCGCTTAGGATTTGGAGAGAAAGCGGTTTGAAGAA	180		
QY	1490	GAGCGAGCAGCTGGGTAAAGCAGCAGTTTTAAACATGACGAACTTTGACCACAGAAC	1549		
DB	181	GAAAGAGCAGTGGTTAAAGCAACAGTTTTTAAACATGACTACCTTTGACCACAGAAC	240		
QY	1550	TCAGAAATGTGAACATTTTCAGTGCCTTCTCAGGAAGTCTCGATCCAGACAAATCTTATA	1609		
DB	241	TCAGAAATATGAACATTTTCAGTGCCTTCTCAGGAAGTCTCGATCCGAGACATCTCTCA	300		
QY	1610	GTCCACTCAGCGCCACGCAAGAGCTACAGTGTGGCTAATGGGGTCCAGCTTGC	1669		
DB	301	GTGCACTCGGGCCACGCAAGAGGCTCACAGTGTGTCTAACGGGACTCCAGTTGC	360		
QY	1670	ACATCAAACTGACTAAATCTTCTCCTCCCTCACCCTTCTACTTCAGACTTTCCGCAGACA	1729		
DB	361	ACCTTAAACTTACTAAGTCTCTTCTCCTGCTTCTCTCTGACTTTCGAGACTTTGCCAGACA	420		
QY	1730	CATTTCATGTGTCTGAACACAGTTTCCATCAGTGTGCTGAATATAACTCTCTGAAGAAAGT	1789		
DB	421	CGGTCTCGGCATCTGACCATAGTTTCCATCAATGTACTGATATATAACTCTCTGAAGAACT	480		
QY	1790	AAACCAAGTAGTTTGAAGAGAAAGCAAGCATGAGAGTGGAGCTGCACTCGAGGCCC	1849		
DB	481	AAACCAATACAGTTCGCGAGAGAGGTACAAATCAGAAATGGAGCATGGCATCAAGACCT	540		
QY	1850	AGCTCGGGGAGGGTGTCTACAGGGGATGCTTCTCGGCTTTCAGAGCGCTCAGGGGAC	1909		
DB	541	GGATCAGAGGAGGGTGTCTACAGTGGATGCTTCTTTCACCTACAAATCCCATGTAGAA	600		
QY	1910	CGAGATGACTTACCTTAAATGTGGGGCTGCAGTGTCTTTC	1950		
DB	601	AAAGATGACTTACCTTAGACGCTCTGGACTGGAAATGTTTTC	641		

RESULT 8	
ID	ABA20636 standard; DNA; 6619 BP.
XX	ABA20636;
XX	ABA20636;
DT	23-JAN-2002 (first entry)
XX	Human nervous system related polynucleotide SEQ ID NO 12967.
DE	Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW	antiparkinsonian; antickling; antianaemic; antiarthritic; cancer;
KW	antiathermatic; hepatotropic; cerebroprotective; antiinflammatory;
KW	antiallergic; antididiabetic; antitumor; anticonvulsant; antifungal;
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
OS	Homo sapiens.
XX	WO200159063-A2.
XX	16-AUG-2001.
XX	17-JAN-2001; 2001WO-US001334.
PF	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.
PR	07-JUN-2000; 2000US-0209467P.
PR	28-JUN-2000; 2000US-0214886P.
PR	30-JUN-2000; 2000US-0215135P.
PR	07-JUL-2000; 2000US-0216647P.
PR	11-JUL-2000; 2000US-0216880P.
PR	11-JUL-2000; 2000US-0217487P.
PR	14-JUL-2000; 2000US-0217496P.
PR	26-JUL-2000; 2000US-0220963P.
PR	26-JUL-2000; 2000US-0220964P.
PR	14-AUG-2000; 2000US-0224518P.
PR	14-AUG-2000; 2000US-0224519P.
PR	14-AUG-2000; 2000US-0225213P.
PR	14-AUG-2000; 2000US-0225214P.
PR	14-AUG-2000; 2000US-0225266P.
PR	14-AUG-2000; 2000US-0225267P.
PR	14-AUG-2000; 2000US-0225268P.
PR	14-AUG-2000; 2000US-0225270P.
PR	14-AUG-2000; 2000US-0225447P.
PR	14-AUG-2000; 2000US-0225757P.
PR	14-AUG-2000; 2000US-0225758P.
PR	18-AUG-2000; 2000US-0225759P.
PR	22-AUG-2000; 2000US-0226681P.
PR	22-AUG-2000; 2000US-0226868P.
PR	23-AUG-2000; 2000US-0227009P.
PR	30-AUG-2000; 2000US-0228924P.
PR	01-SEP-2000; 2000US-0229287P.
PR	01-SEP-2000; 2000US-0229343P.
PR	01-SEP-2000; 2000US-0229344P.
PR	05-SEP-2000; 2000US-0229345P.
PR	05-SEP-2000; 2000US-0229513P.
PR	06-SEP-2000; 2000US-0230437P.
PR	06-SEP-2000; 2000US-0230438P.
PR	08-SEP-2000; 2000US-0231242P.
PR	08-SEP-2000; 2000US-0231243P.

PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 13-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-0242221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.

PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX Disclosure; SEQ ID NO 12967; 1701pp + Sequence Listing; English.  
PS The invention relates to novel genes (ABAI1004-ABA21534) and proteins  
CC (ABBI4678-ABBI8001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
XX  
SQ Sequence 6619 BP; 2137 A; 1013 C; 1235 G; 2234 T; 0 U; 0 Other;  
Query Match 14.9%; Score 399.8; DB 5; Length 6619;  
Best Local Similarity 74.0%; Pred. No. 8.2e-98;  
Matches 738; Conservative 0; Mismatches 202; Indels 57; Gaps 16;  
Qy 1749 ACAGTTCATCAGTGTGCTGAATATAACTCTGAAGAAAGTAAACCAAGTGAGTTGCAA 1808  
Db 2916 ACAGTTCATCAATGTAATCAATATACTGCTGAAGAAATTAACCAAAATCAGGTTGGAG 2975  
Qy 1809 GAGAAACGACGGATCAGAACTGGAGCGTGCAGTCGAGCCGAGTCCGGGAGGGTGCT 1868  
Db 2976 GAGAAATGACAAATCAAAAATGGAGTGTGGCATCAAGACCTGGATCAGAGAGGTTGCT 3035  
Qy 1869 ACAGCGGATGCTCTCGGCGCTTCAGGAGCGCTACGGGGACCGAGATGACTTACTTTAAA 1928  
Db 3036 ATAGTGGATGCTCTCTGAGCTACAAATTTCTCATGTAGAAAAGATGACTTACTTTAGA 3095  
Qy 1929 TGTGCGGGCTGCAGTGTGTTCCAGATGTGCGGTAGAGG-----AGTTGACAC 1977  
Db 3096 CATGTGGACTGGAATTTTTTTCATTATGTGTTTCATCAAGTTTCACATCTAAGTTGAAAC 3155  
Qy 1978 AGGTGTAGCATAAAGTCAGTCGTC-----TAACTTAAGATGCTCAGAGTTGTTGTTG 2032  
Db 3156 AGGTGTGTATAAAGTCAGTTATCTCTAATAACTTAAGATGGTCTGAGTTGTTGTTG 3215

QY	2033	GACCTCGCTGCTCTTCCCCCAAGAGCTGAAATGCT-AAGCTACTTTAAAAAGGATGCAAGC	2091	XX	17-JAN-2001;	2001WO-US001334.
Db	3216			XX	31-JAN-2000;	2000US-0179065P.
QY	2092	TTTGG- - - -TTGTGTGTAGTAACAGAGCCCTGCTCTGTGACTGCAGGAATGCAT-G	2146	PR	04-FEB-2000;	2000US-0180628P.
Db	3276			PR	24-FEB-2000;	2000US-0184664P.
QY	2147	GGCTTTGGATGGAACAGAGCGCTGGAATGA- -TTGCCCTGCGCAGGTACCCGAGAAGAC	2204	PR	02-MAR-2000;	2000US-0186350P.
Db	3336	ATGTTTGGATGGAACCAAGACACTAGACTGAGTTTCTCTTATAGGTATTAATAATAGC	3395	PR	16-MAR-2000;	2000US-0189874P.
QY	2205	ACTTTTAGGACTGGTTCTGTAAACATTAATATCGTCCCAAGTGTGTTGGCATTTGG	2264	PR	17-MAR-2000;	2000US-0190076P.
Db	3396			PR	18-APR-2000;	2000US-0198123P.
QY	2265	AAGTGTAGCCTTTACTTGAATGTATACTGTAGATTTTAAACAAAGCAGGTTCTATATTTA	2324	PR	19-MAY-2000;	2000US-0205515P.
Db	3456	AGTTTGTAGCCTTTACTTGAATGTATACTGTAGATTTTAAACAAAGCAGGTTCTATATTTA	3515	PR	07-JUN-2000;	2000US-0209467P.
QY	2325	TTATGTTTGTAGTGAATTTTGGGATTAACCTCTTTTCATATATGTTTT- - - - -	2367	PR	28-JUN-2000;	2000US-0214886P.
Db	3516	TTATGTTTGTAGTGTG- - - - -	3574	PR	30-JUN-2000;	2000US-0215135P.
QY	2368	----GTGCTGTACATAATATACATGACTATGTTAAGAGGCTTTAAGGTTTAAAAACTT	2423	PR	07-JUL-2000;	2000US-0216647P.
Db	3575	TGTATGTTTGTACATAGATACACATGATTAATGTTTAAAGAGGCTTTAAAGATTTAAAAAGTTT	3634	PR	07-JUL-2000;	2000US-0216880P.
QY	2424	CACACATGCTTGTAGTATAGCAATTTTCATGCCCAATTAATAATGTTTTTCAGTGCGCATGTTT	2483	PR	11-JUL-2000;	2000US-0217487P.
Db	3635	CACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTTGTGGTATTTCTGTT	3692	PR	14-JUL-2000;	2000US-0217496P.
QY	2484	TACAGAGG- -TTAGGACCACTGCCACATGACAGTTAAGACTTTTATTTTAAGCCACTCTGG	2541	PR	26-JUL-2000;	2000US-0218290P.
Db	3693	TACAGATGTTTGTAGGACCATTCATGATTTACATTTAAGA-ATTCCTTTAATACACTCTGG	3751	PR	26-JUL-2000;	2000US-0220963P.
QY	2542	GCNATAAAATTCAAAGCCCTTCAT-AGCTGAGTTC- - -AGATACTAGAACTACTAA	2597	PR	14-AUG-2000;	2000US-0224519P.
Db	3752	GCAATAAATATTGAAGAGTATTTCCATGAAGCTGAGTTCTTTAGATAATCAACACTACTAA	3811	PR	14-AUG-2000;	2000US-0225214P.
QY	2598	COTTCATTTTGTAGATTTTAAAGCATTT- - -GATTTTATTTTATATATGT- - -GAATGTTAT	2655	PR	14-AUG-2000;	2000US-0225267P.
Db	3812	CATTACATTTTGTAGATTTTATGACATTAGATTTTATTTTGTATATGATAGATATTAT	3871	PR	14-AUG-2000;	2000US-0225268P.
QY	2656	AATTTCTAAGAGGAATTTGATTTATGGAGTAATGGGG	2692	PR	14-AUG-2000;	2000US-0225270P.
Db	3872	AATTTTAAAGGACTATTGATGATAGAGTAAGATAGGG	3908	PR	14-AUG-2000;	2000US-0225275P.
RESULT 9						
ABA20638	ID	ABA20638	standard; DNA; 6619 BP.			
XX	AC	ABA20638;				
XX	XX					
DT	23-JAN-2002	(first entry)				
DE	Human nervous system related polynucleotide	SEQ ID NO 12969.				
KW	Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;					
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;					
KW	antiparkinsonian; antitoxic; antianaemic; antithrictic; cancer;					
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;					
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;					
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;					
XX	neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.					
OS	Homo sapiens.					
XX	WO200159063-A2.					
PN						
XX	16-AUG-2001.					
PD						

```
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0242221P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 12969; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (AB14678-AB18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and antagonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemias; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
XX Sequence 6619 BP; 2134 A; 1012 C; 1235 G; 2238 T; 0 U; 0 Other;

Query Match 14.9%; Score 399.8; DB 5; Length 6619;
Best Local Similarity 74.0%; Pred. No. 8.2e-98;
Matches 738; Conservative 0; Mismatches 202; Indels 57; Gaps 16;

QY 1749 ACAGTTCCATCAGTGTGCTGAATATACTCCTCGAAGAAAGTAAACCAAGTGAGTTGCCAA 1808
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1809 GAGAAAGCAGGATCAGAAAGTGAGCGTGCGAGGCCCGCCAGCTCCCGGAGGGTGCT 1868
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2976 GAGAAAGTACAAATCAAAAATGGAGTGTGGCGTCAAGACCTGGATCAAGAGGTTGCT 3035
QY 1869 ACAGCGATGCTCCTCGGCTTCAGAGCGCTCAGGGAGCCGAGATGACTTACCTTAAA 1928
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3036 ATAGTGGATGCTCCTTGAGCTACACAAATCTCATGTAGAAAAGATGACTTACCTTAGA 3095
QY 1929 TGTGCGGGTGCAGTGTGTTCCAGATGTGCGCTAGAGG-----AGTTGACAC 1977
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3096 CATGTGGACTGGAATTTTTCATTATGTGTTTCATCAAGTTCACATCTAAGTTGAAAC 3155
QY 1978 AGGGTGAGCATAAAGTCAGTCGTC-----TAACCTAAGATGCTCAGAGTGTGTTGTTG 2032
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3156 AGGGTGTGCATAAAGTCAGTTATCTCTAATAACTTAAGATGGTCTCAGTGTGTTGTTG 3215
QY 2033 GACTTCGCTGCTTCCCGCAAGAGCTGAAATGCT-AAGCTACTTAAAGATGCAAGC 2091
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3216 GACTTCCTGCTTCCCGCAAGAGTGAATCTTAATCTTAAAGATATATAAG 3275
QY 2092 TTTGG-----TTGTGTGTAGTAACAGAGCCCTGGCTGTGTGACTCAGAAATGCAT-G 2146
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3276 CTTTGGATATGATTTTGTAGTAACAGAGCATCTGTTCTGTGAATAAAGAAATGTATAG 3335
QY 2147 GCGTTTGGATGGAACAGAGCGCTGGAATGA--TTGCTCGCCAGGTACCGAGAAGC 2204
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3336 ATGTTTGGATGGAACAAAAGCAGTACAGCTGAGTTTCTTTATAGGTATTAAAGATAGC 3395
QY 2205 ACTTTTAGGACTGGTTCTCTGTAAACATTAATAATTCGTCCCAAGTGTGGTTGGCATTGG 2264
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3396 ACTTTTAGGAACGATATTATGTAATGTTTAAATTTTGTCTCAATATAGTTGCAATGG 3455
QY 2265 AAGTGTAGCCTTTACTTTGAATGATATCTGTAGATTTTAAACAGAGGTTCTATATTTA 2324
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3456 AAGTTTAGCCTTTACTTGAATGTACTGTAGATTTTAAACAGAGCGAGTTCTATATTTA 3515
QY 2325 TTATGTTTAGTGATTTTGGGATTAACCTCTCTTCATATGTTT----- 2367
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3516 TTATGTTTAGTG-GTTTGAAGATTAACCTCTCTTCATATGTTTAAATAAAGTGAATTTA 3574
```

QY 2368 ----GTGCTGTACATAAATATACATGACTATGTTAAAGGCTTTAAAGCTTTTAAAAACTT 2423  
Db 3575 TGTATGTTTGTACATAGATACACATGATGTTTAAAGAGGCTTTAAAGATTTTAAAAAGTTT 3634  
QY 2424 CACACCATGCTTGAGTATAGCATTTTCATGCCAATTAAAAATGTTTTCAGTGGCATGGTGT 2483  
Db 3635 CACA-CAACCATAAAGTATAGTATTTTCATGCC-AGTAAAAATTTTTCAGTGGTATTCGTGT 3692  
QY 2484 TACAG--AGTTAGGACCTGCCACATGACAGTTTAAAGCTTTTATTTTAAAGCCACTCTGG 2541  
Db 3693 TACAGATATATTAGGACCATTCATGATGATTTACATTTAAGA-ATTCTCTTTAATACATCTGG 3751  
QY 2542 GCAATAAAATTCAAAGCCCTTCAT-AGCTGAGTTC---AGATAACTAGAACTACTAA 2597  
Db 3752 GCAATAAATATTGAAGGTAATTCATGAAGCTGAGTCTTTTAGATAAATCAACACTACTAA 3811  
QY 2598 CCGTACATTTTGGAGATTTTAAAGCAAT-GTATTTTATTTTATATATGTT-GAATGTTAT 2655  
Db 3812 CATTACATTTTGGAGATTTTATGACATTAGATTTTATTTTGTATATGTAGAAATATTAT 3871  
QY 2656 AATTCTTAAGAGAAATATTGATTATGGAGTAATGGG 2692  
Db 3872 AATTTTAAAGGACTATTGTATGATAGAGAATAGGG 3908

## RESULT 10

ID ABA20637  
XX ABA20637 standard; DNA; 6619 BP.  
XX AC ABA20637;  
XX XX  
DT 23-JAN-2002. (first entry)  
XX XX  
DE Human nervous system related polynucleotide SEQ ID NO 12968.  
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antistickling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX Homo sapiens.  
OS  
XX  
XX  
FN WO200159063-A2.  
XX  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234987P.  
PR 25-SEP-2000; 2000US-0234988P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 20-OCT-2000; 2000US-024221P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.

PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246529P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 01-DEC-2000; 2000US-0251160P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
XX Disclosure; SEQ ID NO 12968; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA215134) and proteins  
CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 6619 BP; 2137 A; 1013 C; 1233 G; 2236 T; 0 U; 0 Other;  
Query Match 14.9%; Score 399.8; DB 5; Length 6619;  
Best Local Similarity 74.0%; Pred. No. 8.2e-98;  
Matches 738; Conservative 0; Mismatches 202; Indels 57; Gaps 16;  
QY 1749 ACAAGTTCCATCAGTGTGCTGAATATTAACCTCCTGAAGAAAGTAAACCAAGTGAAGTTCGCAA 1808  
DB 2916 ACAGTTCAATCAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2975  
QY 1809 GAGAAAGCAGGATCAGAAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1868  
DB 2976 GAGAAATGTACAAATCAAAATGAGTGTGGCATCAAGACCTGATGATCAAGAGAGGTTGCT 3035  
QY 1869 ACAGCGATGCTCCTCGGCTTCAGGAGCGCTCAGGGGACCGAGATGACTTACCTTAAA 1928  
DB 3036 ATAGTGGATGCTCCTTGGAGCTACACAAATTCCTATGTAGAAAAGATGACTTACCTTAGA 3095  
QY 1929 TGTGGGGCTGCAGTGTCTCCAGATGTCGCTAGAGG-----AGTTGACAC 1977  
DB 3096 CATGTGGACTGGAATTTTTTTCATTAATGTGTTTCATCAAGTTTCACATCTAAGTTGAAAC 3155  
QY 1978 AGGGTGTAGATAAAGTCACTCGTC-----TAACTTAAAGATGCTCAGAGTTGTTTGTG 2032  
DB 3156 AGGGTGTGTATAAAGTCACTTATCTCTAATAAGTCACTTAAAGTGTGCTGAGTTGTTGTTG 3215  
QY 2033 GACTTCGCTGCTCTCCGCTCCCAAGAGCTGAATGCT-NAAGTACTTAAAGGATGCAAGC 2091  
DB 3216 GACTTCCTGCTCTCCGCTCCCAAGAGTGAATCTTAAATCTATTTAAAGGATATAAAG 3275  
QY 2092 TTTGG-----TTGTGTGTAGTAACAGAGCCCTGCTCTGTGACTCGAGGAATGCAAT-G 2146  
DB 3276 CTTTGGATATGATTTTGTAGTACAGNAGCATCTGGTTCCTGTGAATAAGGAATGTATAG 3335  
QY 2147 GCGTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCCAGGTACCGAGAGAGC 2204  
DB 3336 ATGTTGGATGGAACACAAAGCACATAGACTGAGTTTCCTCTATATAGGTATTTAAATAATAGC 3395  
QY 2205 ACTTTAGGACTGGTTCCTGTAAACATTAAATATTCGTCCTCCCAAGTCTGGTTGGCATTGG 2264  
DB 3396 ACTTTAGGAACTGATTTTGTAAATGTTTAAATTTTGTCTCAAAATATAGTTGGCATTGG 3455  
QY 2265 AAGTGTAGCCTTTACTTTGAATGTACTGTAGATTTTAAACAAAGCAGGTCTTATATTTA 2324  
DB 3456 AAGTTTAGCCTTTACTTTGAATGTACTGTAGATTTTAAACAAAGCAGGTCTTATATTTA 3515  
QY 2325 TTATGTTTAGTGTGATTTTGGGATTAACCTCTTTTCATATGTTTT----- 2367  
DB 3516 TTATGTTTAGTGTG-GTTTGAATTTACCTCTTTTCATATGTTTTTAAATPAAAGTGAATTTA 3574  
QY 2368 ----GTGCTGTACATAAATATACATGACTATGTTTAAAGGCTTTAAAGGTTTAAAGACTT 2423  
DB 3575 TGTATGTTTTGTACATAGATACACATGATTTATGTTAAAGGCTTTTAAGATTTTAAAGTTT 3634  
QY 2424 CACACCATGCTTGAGTATAGCATTTTCATGCCAATTTAAATTTTTCAGTGGCATGGTGT 2483  
DB 3635 CACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTCAGTGGTATCTGTT 3692  
QY 2484 TACAGAG--TTAGGACCATGCCCATGACATGACATGATTAAGACTTTATTTTAAAGCCTCTGG 2541  
DB 3693 TACAGATGTTATAGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3751  
QY 2542 GCATTAATAATTCAAAGCCCTTCAT--AAGCTGAGTTTC--AGATACTAGAACTACTAA 2597  
DB 3752 GCATTAATAATTCAAAGGATTTTCATGAAAGCTGATTTTCATGAAAGCTGATTTTCATGAA 3811  
QY 2598 CGTTACATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTT 2655  
DB 3812 CATTACATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTT 3871  
QY 2656 AATTCTTAAGAGGAATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTTTTCAGATTT 2692  
DB 3872 AATTTTAAAGGAGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3908







QY 1095 TGACAAACAGCATCAGGAACACAGTGGAGAAATTTGAAAGATCATGTAGAAAACTCGATA 1154  
 Db 358 TTACAAACAGCATCAGAAACACAGTGGAGAAATTTGAAAGATCATGTAGAAAACTCGATA 417  
 QY 1155 ACCAAGCTTCGAGGTACACTCAGAGGGCTTAAATGAGGAGGACGTCATCTCACGACAAG 1214  
 Db 418 ACCAAGCTTCGAGGTACACTCAGAGGGCTTAAATGAGGAGGACGTCATCTCACGACAAG 477  
 QY 1215 ACCATGAGCAAGAGACTGAGA 1235  
 Db 478 ACCATGAGCAAGAGAACTGAAA 498

## RESULT 14

AAAF66702

ID AAF66702 standard; cDNA; 403 BP.

AC AAF66702;

XX 09-APR-2001 (first entry)

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2458.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

KW breast cancer; lung cancer; cancer detection; ss.

KW Homo sapiens.

OS WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US018374.

XX 02-JUL-1999; 99US-0142310P.

PR 02-JUL-1999; 99US-0142311P.

XX (CHIR ) CHIRON CORP.

PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;

PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;

PI Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;

XX WPI; 2001-091805/10.

DR Library of polynucleotides for diagnosing a cancerous state of a

XX mammalian cell and detecting cancer, particularly of the colon or

PT prostate, comprises 3351 human polynucleotide sequences.

XX Claim 9; Page 906; 1046pp; English.

XX The present sequence is one of 3351 sequences in a library of human

CC polynucleotides. The library is used to detect differentially expressed

CC genes correlated with a cancerous state of a mammalian cell and can

CC detect colon, prostate, breast and lung cancer. The library can be used

CC to produce probes for detection of mRNA and to produce additional copies

CC of the polynucleotides. The probes can be used for chromosome mapping of

CC the polynucleotide and for detection of transcription levels. Ribozymes

CC or antisense oligonucleotides can be generated. The polynucleotides and

CC their gene products are used as genetic or biochemical markers (e.g. in

CC blood or tissues) that will detect the earliest changes along the

CC carcinogenesis pathway and/or monitor the efficacy of therapies and

CC preventive interventions. The polynucleotides, polypeptides and

CC antibodies against them can be used in pharmaceutical compositions to

CC treat the cancers and proliferative disorders such as neoplasia,

CC dysplasia and hyperplasia

XX Sequence 403 BP; 162 A; 57 C; 96 G; 88 T; 0 U; 0 Other;

SQ Query Match 11.4%; Score 305.8; DB 5; Length 403;

XX Best Local Similarity 85.6%; Pred. No. 9e-73;

Matches 340; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
 QY 470 ATGACACACCTGCAGAGCTGTACGCCAAACTTAAGAGGACGCTTGGAAACGTCACGCGG 529  
 Db 7 AGGGACCATCTACAGAGCTGTCTCAAACTTATGGAACAACCTGGAACCTCCAGGAG 66  
 QY 530 GAGATGATCGGGCTTCAAGAGAGAGACAGGAGCTGCAGTGCAAGAACAGGAGTTTGCAT 589  
 Db 67 GAAATGATTGGGCTTCAGGAAAGAGACAGACAGTTCACAATGTAAAGACAGGAATTTGCAT 126  
 QY 590 CAGCTCCTGGAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCCG 649  
 Db 127 CAGCTACTAAAGATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCATAGCCAG 186  
 QY 650 GCTACTCAGTATAATCATGTGTGAAGAGGAGCGTGAATATAAATGCTTAAAGGAG 709  
 Db 187 GCTACTCAGTATAATCATGTGTGAAGAGGAGCGTGAATATAAATGCTTAAAGGAG 246  
 QY 710 GCTCTGATCAGCTGTTATGAACAAGAGGATATAAATCATAGCCAGTATTTTAAAT 769  
 Db 247 CGTCTACATCAACTTGTATGAACAAGAGGATATAAATCATAGCCAGTATTTTAAAT 306  
 QY 770 TATGTGGCTCGAGCTGTGCAACGAGGCTCATGAGGAGCTGCACAAAACAGACCCAGG 829  
 Db 307 TATGTGGGAGAGCTGTGGAAGAGGCTCCTGAGGAGCTGTTAAACCTGAAGCCAGG 366  
 QY 830 AATGAAGATGAGATGTACAAAATTTCTGTGAATGATT 866  
 Db 367 AATGAAGATGAGATGTATATAAATTTCTTGAATGATT 403

## RESULT 15

AAC03006

ID AAC03006 standard; cDNA; 428 BP.

AC AAC03006;

XX 06-OCT-2000 (first entry)

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 3004.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

OS EP1033401-A2.

PN 06-SEP-2000.

PD 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG03000.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

CC obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

CC diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX Claim 1; SEQ ID NO 3004; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC	well suited for isolating cDNA sequences derived from the 5' ends of
CC	mRNAs and even in those cases where longer cDNA sequences have been
CC	obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC	mRNAs with intact 5' ends and can therefore be used to obtain full length
CC	cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC	gene therapy and chromosome mapping procedures. They are used to obtain
CC	upstream regulatory sequences and to design expression and secretion
CC	vectors
XX	
SQ	Sequence 428 BP; 170 A; 69 C; 94 G; 95 T; 0 U; 0 Other;
	Query Match            11.0%; Score 297.2; DB 3; Length 428;
	Best Local Similarity   82.3%; Pred. No. 2.1e-70;
	Matches 354; Conservative 0; Mismatches 73; Indels 3; Gaps 1
QY	878 CAGAAGCAGATCCTGTGGAGAACGCGGAGCTGAAGAAGGTCTCCAGCAGATGAAGAAG 937
DB	
DB	1 CAGAAACAAATCCTTANTGGAANAATGCAGACTTAAGAAGGTTCCTCAACAATGAAAAG 60
QY	938 GAGATGATCTCTCTCTGTCTCTCTCAGAAGAAGAACGCCAGGAAAGACGAGACGCGC 997
DB	
DB	61 GAAATGATTCTCTCTCTCTCTCCCCAAAAAGAAACCTCAGAGAAAGAGTAGATGATAGT 120
QY	998 ACAGGCACCTGTTGCTATCTCCGATATAGAAGATGACTCTGGGGAACCTCAGCAGACACAGC 105
DB	
DB	121 ACAGGAACCTGT---TATTTCGATGTTGAAGAAGATGCCGGGAACTAAGCAGAGAGAGT 177
QY	1058 GTGTGGGGCCCTTCTCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAAACAG 111
DB	
DB	178 ATGTGGGACCTTTTCTGTGAACTGTGAGAGAGCAGCTTACAAACAGCATCNGAANAACAG 237
QY	1118 TGGAGAAATTTTTGAAAGTCATGTAGAAAAAATCTCGATAACCAGCTTCGAAGGTACACTCA 1177
DB	
DB	238 TGGAGAAATTTTTGAAAGTCATGTAGAAAAGCTTGATAACCAAGTTTCAAAGGTACACCTG 297
QY	1178 GAGGGCCTTAATGAGGAGGAGCTGCTCTCACGAAGAACCATGACGACAGAGNCTGAGAA 123
DB	
DB	298 GAAGGTTTTTAATGATGAAGATGTTAATCTTCAGCAAGAAGCCATGAACAAGAACTGAAAA 357
QY	1238 CTGAGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAG 129
DB	
DB	358 CTCAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAANAATCAGCAACAGCTTTTACAG 417
QY	1298 CAGCAGCTGG 1307
DB	
DB	418 CAGCAGCTCG 427

Search completed: June 13, 2006, 20:46:44  
Job time : 1470 secs



Query Match 100.0%; Score 3165; DB 8; Length 615;  
Best Local Similarity 100.0%; Pred. No. 5.2e-215; Mismatches 0; Indels 0; Gaps 0;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGDWMTVDPVLCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVGFVCTG 60  
Db 1 MGDWMTVDPVLCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVGFVCTG 60

Qy 61 NTEQISYLDQELTTFGPPSLYEESKSKAEKRELINIVAVLNCMNEELLVLRKNLLAQESV 120  
Db 61 NTEQISYLDQELTTFGPPSLYEESKSKAEKRELINIVAVLNCMNEELLVLRKNLLAQESV 120

Qy 121 ETQNLKLGSDMDHLSQYAKLEQLETSRRMIGLQERDRQLQCKNRSIHLQKLNKDE 180  
Db 121 ETQNLKLGSDMDHLSQYAKLEQLETSRRMIGLQERDRQLQCKNRSIHLQKLNKDE 180

Qy 181 OKLQNIISRATQYNDVKKREYNKLERLHQLVMNKKDNIAAMDVLNYYVGRADGKRG 240  
Db 181 OKLQNIISRATQYNDVKKREYNKLERLHQLVMNKKDNIAAMDVLNYYVGRADGKRG 240

Qy 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMKEMISLLSPQKKPR 300  
Db 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMKEMISLLSPQKKPR 300

Qy 301 ERAEDGTGTVVAISDIEDSGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHVEKLDNQ 360  
Db 301 ERAEDGTGTVVAISDIEDSGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHVEKLDNQ 360

Qy 361 ASKVHSEGLNEEDVISRQDHEQTEKLELEIERCKEMIKAAQQLLQQQLATTCDDDTTSL 420  
Db 361 ASKVHSEGLNEEDVISRQDHEQTEKLELEIERCKEMIKAAQQLLQQQLATTCDDDTTSL 420

Qy 421 LRDCYLLEEKERLKEEWTLFKQKKNFERERSFTEAAIRGLERKAFEEERASVWKQF 480  
Db 421 LRDCYLLEEKERLKEEWTLFKQKKNFERERSFTEAAIRGLERKAFEEERASVWKQF 480

Qy 481 LNMNFDHQNSENVKLFSAFSGSSDPDNLIVHSRPRQKHLHSHVANGVPACTSKLTKSLPA 540  
Db 481 LNMNFDHQNSENVKLFSAFSGSSDPDNLIVHSRPRQKHLHSHVANGVPACTSKLTKSLPA 540

Qy 541 SPSTSDFRQTHSCVSEHSSISVLNITPESKPSVARESTDQKWSVQSRPSREGCYSGC 600  
Db 541 SPSTSDFRQTHSCVSEHSSISVLNITPESKPSVARESTDQKWSVQSRPSREGCYSGC 600

Qy 601 SSAFRSAHGDRDDL 615  
Db 601 SSAFRSAHGDRDDL 615

RESULT 2  
ADO26493  
ID ADO26493 standard; protein; 613 AA.  
XX ADO26493;  
AC ADO26493;  
CC ADO26493;  
DT 29-JUL-2004 (first entry)  
XX DE Rat afadin dilution domain binding protein (ADIP) SeqID 4.  
XX  
KW rat; afadin dilution domain binding protein; ADIP; afadin; actinin;  
KW binding inhibitor; cardiant; heat disease; myocardial infarction;  
KW myocarditis.  
XX  
OS Rattus norvegicus.  
XX  
XX JP2004135658-A.  
XX  
XX 13-MAY-2004.  
XX  
XX 14-AUG-2003; 2003JP-00293554.  
XX  
XX 27-SEP-2002; 2002JP-00284263.

XX (EISA ) EISAI CO LTD.  
XX PA  
XX WPI; 2004-404616/38.  
DR N-PSDB; ADO26492.  
XX  
PT New polynucleotide encoding an afadin dilution domain binding protein  
PT having avidity with afadin or actinin, useful for diagnosing heart  
PT diseases e.g. myocardial infarction.  
XX  
PS Claim 1; SEQ ID NO 4; 37pp; Japanese.  
XX  
CC This invention relates to a novel isolated nucleic acid encoding an  
CC afadin dilution domain binding protein (ADIP) that exhibits an avidity  
CC with afadin/actinin. Specifically, it refers to screening assays to  
CC identify compounds that modulate ADIP avidity and provides suitable  
CC agonists, antagonists and antibodies thereof. The present invention  
CC provides methods to identify afadin and actinin binding inhibitors  
CC therapeutically as cardiants to diagnose and/ or treat heart disease such  
CC as myocardial infarction or myocarditis. This polypeptide sequence is the  
CC rat ADIP protein sequence of the invention.  
XX  
SQ Sequence 613 AA;

Query Match 92.8%; Score 2937; DB 8; Length 613;  
Best Local Similarity 92.2%; Pred. No. 6.9e-199;  
Matches 567; Conservative 18; Mismatches 28; Indels 2; Gaps 1;

Qy 1 MGDWMTVDPVLCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVGFVCTG 60  
Db 1 MGDWMTVDPVLCTENKNLSQYTSKMPSSLYSQVLCSSVPLSKNVHGVGFVCTG 60

Qy 61 NIEQISYLDQELTTFGPPSLYEESKSKAEKRELINIVAVLNCMNEELLVLRKNLLAQESV 120  
Db 61 NIEQISYLDQELTTFGPPSLYEESKSKAEKRELINIVAVLNCMNEELLVLRKNLLAQESV 120

Qy 121 ETQNLKLGSDMDHLSQYAKLEQLETSRRMIGLQERDRQLQCKNRSIHLQKLNKDE 180  
Db 121 ETQNLKLGSDMDHLSQYAKLEQLETSRRMIGLQERDRQLQCKNRSIHLQKLNKDE 180

Qy 181 OKLQNIISRATQYNDVKKREYNKLERLHQLVMNKKDNIAAMDVLNYYVGRADGKRG 240  
Db 181 OKLQNIISRATQYNDVKKREYNKLERLHQLVMNKKDNIAAMDVLNYYVGRADGKRG 240

Qy 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMKEMISLLSPQKKPR 300  
Db 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMKEMISLLSPQKKPR 300

Qy 301 ERAEDGTGTVVAISDIEDSGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHVEKLDNQ 360  
Db 301 ERAEDGTGTVVAISDIEDSGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHVEKLDNQ 360

Qy 361 ASKVHSEGLNEEDVISRQDHEQTEKLELEIERCKEMIKAAQQLLQQQLATTCDDDTTSL 420  
Db 361 ASKVHSEGLNEEDVISRQDHEQTEKLELEIERCKEMIKAAQQLLQQQLATTCDDDTTSL 420

Qy 421 LRDCYLLEEKERLKEEWTLFKQKKNFERERSFTEAAIRGLERKAFEEERASVWKQF 480  
Db 421 LRDCYLLEEKERLKEEWTLFKQKKNFERERSFTEAAIRGLERKAFEEERASVWKQF 480

Qy 481 LNMNFDHQNSENVKLFSAFSGSSDPDNLIVHSRPRQKHLHSHVANGVPACTSKLTKSLPA 540  
Db 481 LNMNFDHQNSENVKLFSAFSGSSDPDNLIVHSRPRQKHLHSHVANGVPACTSKLTKSLPA 540

Qy 541 SPSTSDFRQTHSCVSEHSSISVLNITPESKPSVARESTDQKWSVQSRPSREGCYSGC 600  
Db 541 SP - -SDFCPSRSCVSEHSPVSALTVTPEETKPNVGRSESTDQKWSVQSRPSREGCYSGC 598

Qy 601 SSAFRSAHGDRDDL 615  
Db 599 SSATTSSEHVERDDL 613

RESULT 3  
 ABO14751  
 ID ABO14751 standard; protein; 567 AA.  
 XX  
 AC ABO14751;  
 XX  
 DT 25-AUG-2003 (first entry)  
 XX  
 DE Novel human protein #124.  
 XX  
 KW Human; NOV; gene therapy; endocrine related disease; diabetes;  
 KW metabolism-related disease; obesity; central nervous system disorder;  
 KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;  
 KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;  
 KW psoriasis; allergy; lupus erythematosus; asthma; cancer;  
 KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;  
 KW colon cancer; liver cancer; breast cancer; ovarian cancer;  
 KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;  
 KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;  
 KW stroke; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003023002-A2.  
 XX  
 PD 20-MAR-2003.  
 XX  
 PF 09-SEP-2002; 2002WO-US028539.  
 XX  
 PR 07-SEP-2001; 2001US-0318120P.  
 PR 07-SEP-2001; 2001US-0318130P.  
 PR 10-SEP-2001; 2001US-0318430P.  
 PR 17-SEP-2001; 2001US-0322636P.  
 PR 17-SEP-2001; 2001US-0322781P.  
 PR 17-SEP-2001; 2001US-0322816P.  
 PR 17-SEP-2001; 2001US-0322817P.  
 PR 19-SEP-2001; 2001US-0323519P.  
 PR 20-SEP-2001; 2001US-0323631P.  
 PR 20-SEP-2001; 2001US-0323636P.  
 PR 25-SEP-2001; 2001US-0324969P.  
 PR 26-SEP-2001; 2001US-0324990P.  
 PR 17-APR-2002; 2002US-0373212P.  
 PR 06-SEP-2002; 2002US-00236177.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Spytek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;  
 PI Gerlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;  
 PI Shinkets RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;  
 PI Rieger DK, Taupier RU, Shenoy SG, Liu X, Padigar M, Alsobrook JP;  
 PI Lepley DM, Edinger SK, Burgess CE;  
 XX  
 WPI; 2003-313242/30.  
 DR N-PSDB; ACD19444.  
 XX  
 PT New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOV)  
 PT and polynucleotides, useful in gene therapy, e.g. for treating or  
 PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,  
 PT stroke or infections.  
 XX  
 PS Claim 1; Page 353; 586pp; English.  
 XX  
 CC The invention describes a new isolated polypeptide (NOV). The NOV  
 CC polypeptide, nucleic acid and antibody are useful as therapeutics,  
 CC particularly in the manufacture of a medicament for treating a syndrome  
 CC associated with a human disease, which includes a pathology associated  
 CC with NOV polypeptide. The DNA encoding the protein is useful in gene  
 CC therapy for treating the disease or condition. In particular, the NOV  
 CC polypeptide or polynucleotide is useful for treating endocrine/  
 CC metabolism-related diseases (e.g. obesity or diabetes), central nervous  
 CC system disorders (e.g. Alzheimer's disease, Parkinson's disease,  
 CC epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune

CC and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,  
 CC asthma, inflammatory bowel disease, rheumatoid arthritis or  
 CC osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,  
 CC prostate or brain cancers, or melanoma), liver diseases (e.g. liver  
 CC cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),  
 CC haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).  
 CC These are also useful in developing powerful assay system for functional  
 CC analysis of various human disorders, as well as in diagnostic  
 CC applications, and for monitoring the effects of drugs during clinical  
 CC trials. This is the amino acid sequence of a novel human NOV protein  
 XX  
 SQ Sequence 567 AA;  
 Query Match 82.1%; Score 2598.5; DB 6; Length 567;  
 Best Local Similarity 89.1%; Pred. No. 5.5e-175;  
 Matches 506; Conservative 26; Mismatches 35; Indels 1; Gaps 1;  
 QY 28 MPSSLYSOOVLCSYVPLSKNVHGVFCVCTGNIQSISYLDDELTTGFPSPSYESKS 87  
 DB 1 MPSSLYSOOVLCSYVPLSKNVHGVFCVCTGNIQSISYLDDELTTGFPSPSYESKG 60  
 QY 88 KEAKRELNIIVAVLNCNELLVLRKNLLAQESVETQNLKLGSDMDHLQSCYAKLKEOLET 147  
 DB 61 KETRELNIIVAVLNCNELLVLRKNLLAQENVETQNLKLGSDMDHLQSCYKLEOLET 120  
 QY 148 SRREMIGLQERDRQLQCKNRSLHQLKNEKDEYQKLNIIASRATQYNHDKKEREYNK 207  
 DB 121 SRREMIGLQERDRQLQCKNRSLHQLKNEKDEYQKLNIIASRATQYNHDKKEREYNK 180  
 QY 208 LKERLHQLVNNKDKNIAMDVINYVGRADGKGSWRTDKTEARNEDEMYKILLNDYEQ 267  
 DB 181 LKERLHQLVNNKDKKIAMDVINYVGRADGKGSWRTDKTEARNEDEMYKILLNDYEQ 240  
 QY 268 KOILMENAEELKVLQOMKEMISLLSPQKKPRERAEDGTGTVAISDIEDDSGELSDSV 327  
 DB 241 KOILMENAEELKVLQOMKEMISLLSPQKKPRERVDSDGTGTV-ISDVEDAGELSDSV 299  
 QY 328 WGLSCDTVREQLTNSIRKQWIRLKSHEKLDNOASKVHSEGLNEEDVI SRQDHEQTEKL 387  
 DB 300 WDLSCETVREQLTNSIRKQWIRLKSHEKLDNOASKVHSEGLNEEDVI SRQDHEQTEKL 359  
 QY 388 ELEIERCKEMIRAKQOLLOQLATTCDDDTTSLRDCYLLEKERLKEWTLPEKOKNF 447  
 DB 360 ELEIOCKEMIRAKQOLLOQLATYDDDTTSLRDCYLLEKERLKEWTLPEKOKNF 419  
 QY 448 ERERSFTAAIRLGLERKAFEEERASWVKQFQFLNMTNFDHONSENKVLFSAPSGSSDPD 507  
 DB 420 ERERSFTAAIRLGLERKAFEEERASWVKQFQFLNMTTDFHONSENKVLFSAPSGSSDWD 479  
 QY 508 NLIVHSRPPQKXKLVANGVPACTSKLTKSLPASSTSDPQTHSCVSEHSSISVLNITP 567  
 DB 480 NLIVHSRPPQKXKLVANGVPACTSKLTKSLPASSTSDPQTHSCVSEHSSISVLNITP 539  
 QY 568 EESKPSSEVARESTDQKWSVQSRPSSREG 595  
 DB 540 EBKPNQVGERTNQKWSVASRPGSQEG 567  
 RESULT 4  
 ADB65313  
 ID ADB65313 standard; protein; 504 AA.  
 XX  
 AC ADB65313;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Human protein encoded by clone TESTI20071130.  
 XX  
 KW Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;  
 KW cell regeneration; membrane protein; signal transduction-related protein;  
 KW transcription-related protein; osteoporosis; neurological disease;  
 KW cancer; tumour.  
 XX







PA (TABA/) TABASKA J E.  
XX (CAOV/) CAO Y.  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
XX Claim 1; SEQ ID NO 54149; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
XX invention.  
XX  
XX Sequence 443 AA;  
SQ  
Query Match 12.0%; Score 379.5; DB 8; Length 443;  
Best Local Similarity 26.6%; Pred. No. 3.5e-18;  
Matches 106; Conservative 101; Mismatches 135; Indels 57; Gaps 14;  
QY 27 KMSPSLY-----SQVLCSYVPLSKNVHGVFCTGENIEQISYILDQELTTGFPSSL 81  
DB 54 RMSSSARFDRASSQOL--QPPPSHASMSDDGAFANAENLHCARYLNQTLVTFGPPA- 110  
QY 82 YEESKSKEAKRLNTAVLNCNELLVLQKNLLAQESVETONKLGSDMDHLQSCYAKL 141  
DB 111 ---SIDLFTDVPSTARTCNLCIYALQQRDIREFRESTNDQRQMSDISLEAKIERM 167  
QY 142 KQLETSRREMIGLQERDRLOQCKNRSLHQLLNKEDEVQKLNITIASR--ATQYNHDVK 199  
DB 168 DAQLAAKDELATLTRTEAKNTAALKSQIDKLQERDEFQKM--VIGNQVRTQIHEMK 225  
QY 200 RKEREYNKLERLHQLVNMKKDN--IAMDVLNYVGRADKKGKSWRTDKTEARNEDEMYK 257  
DB 226 KKEKEYIKLQELKNQVLMKEKKESRSRGMEINLLQKEGRQGTWNGKK---NDNDYYK 281  
QY 258 ILLNDYEVROKILMENAEKVKVLOQMKEMISLLSPQKKPRERAEDGTGTVATSDI-E 316  
DB 282 MIVDAYEVKKQELMGENADRLALLSMQMDRDFLN-----APNGSSQSTVTDNGR 332  
QY 317 DDSGEL-----SRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDN--QASKVHSEGL 369  
DB 333 QESGSPQSLGGKTDVDFLPPHWARDQIBESLRTKMTSIKARMTQLQDAKGAETVSEAT 392  
QY 370 NEEDVISRQDHQETEKLELETERCKEMIKAAQQQLLQQQ 408  
DB 393 DRE-----LELE-----AQLVEA-RSIIQEQ 412  
RESULT 8  
ID AAG40470 standard; protein; 382 AA.  
XX  
AC AAG40470;

XX 18-OCT-2000 (first entry)  
DT  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 50219.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
PD  
XX 25-FEB-2000; 2000EP-00301439.  
PF  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.





```

PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148311P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 9.8%; Score 310; DB 3; Length 347;
Best Local Similarity 24.9%; Pred. No. 2,le-13;
Matches 89; Conservative 89; Mismatches 125; Indels 54; Gaps 11;

QY 56 FCTGENIEOSISYLDQELTTFGFPSPSYBESKSKAKRELNIIVAVLNCMELLVLQRKLL 115
DB 24 FANEDNLEHCTKYLNTQMTVTFGPPA-----SLDLFNDPVSISRTCNMYSLQQRORDIE 79
QY 116 AOSVETQNLKLGSDMDHLSQVAKLKEOLETSSREMIQLQSRDRLOLCKNSLHOLLKN 175
DB 80 FRESANELRQROQSDIARLEAKVERLEALLQOKDREIATITTEAKNTAALKSQIEKLOQ 139
QY 176 EKDEVOKLQNIITASR--ATQYNHDKKEREYNKLERLHQLVM--NKDKNIAMDVLNVY 232
DB 140 ERDEFQRM--VIGNQVKAQQIHEMKKKEKDYIKQLERLNVLMKKEKSRSGMEIMNLL 197
QY 233 GRADGKRGSWRTDKTEARNEDEMYKILLNDIYRQKQIILMENAELKKVLOQMKEMISIL 292
DB 198 QKEGRQRTGWTNGKXTDT-----DFYKKIVDAYEAKNQELMAENTSRLALLRSNQTDMRDL 253
QY 293 SPQKKPRERAEDGTGTVAISDIEDSGELSDSDVWGLSCDVTREQLTNSIRKOWILKS 352
DB 254 N-----APNGSATLAGS--EKREADPSOSPIGGKTESMV--QLQDAPKR----- 293
QY 353 HVEKLDNQASKVHSEGLNEEDVISQDHEQTEKLELEIERCKEMIKAQOQLLQOOL 409
DB 294 -----ASVTSEAT-----ERELE-LEAQLVEARSIIQEESIMSKHL 329

RESULT 10
ADX76571
ID ADX76571 standard; protein; 290 AA.
XX AC ADX76571;
XX DT 23-MAR-2006 (revised)
XX DT 21-APR-2005 (first entry)
XX DE Plant full length insert polypeptide seqid 45937.
XX KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
XX OS Glycine max.
XX US2004034888-A1.
XX PN 19-FEB-2004.
XX PD 28-APR-2003; 2003US-00425114.
XX PF 06-MAY-1999; 99US-00304517.
XX PR 05-NOV-2001; 2001US-00985678.
XX (LIUJ/) LIU J.
XX (ZHOU/) ZHOU Y.
XX (KOVA/) KOVALIC D K.
XX (SCRE/) SCREEN S E.
XX (TABR/) TABASKA J E.
XX (CAOY/) CAO Y.
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX DR
XX PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX PS Claim 1; SEQ ID NO 45937; 15pp; English.

```



PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	16-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144684P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.
PR	22-JUL-1999;	99US-0145089P.
PR	22-JUL-1999;	99US-0145192P.
PR	23-JUL-1999;	99US-0145145P.
PR	23-JUL-1999;	99US-0145218P.
PR	23-JUL-1999;	99US-0145224P.
PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147493P.
PR	09-AUG-1999;	99US-0147935P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 8.8%; Score 277; DB 3; Length 291;		
Best Local Similarity 25.7%; Pred. No. 3 6e-11;		
Matches 75; Conservative 81; Mismatches 90; Indels 46; Gaps 11;		
QY	129	SDMHLCQCYAKLEQLETSRREMIGLQERDRLOQCKNRSLLHOLLKNEKDEVKLQNIITA 188
Db	3	SDISLEAKIERMDAQLAAKORELATLTRXEAKNATAALKSQIDKLQQRDERFQKM--VIG 60
QY	189	SR--ATQYNHDVKRERENKLERHLQVNNKKDKN--IANDVLNVGRADGKGSWRT 244
Db	61	NOQVTFQIHEMKKEKEYIKLOEKLNVLMKEKSSSGMEIWNLLQKRGQRTWNG 120
QY	245	DKTEARNEDEMYKILLNDYEVROKQILMENAEKLVLOQMKEMISLLSPQKKKPRERAE 304
Db	121	KK----NDNDYKMIVDAYEVKKQELMOENADLRALLRSQMOMDRDFLN-----AP 167
QY	305	DGTGTVATSDI-EDDSGEL-----SRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLD 358
Db	168	NGSSQSTVTDNGRQESSGPSPLGGKTDVDFLPFHWARDQIESLRTKMTSIKARMTQLQ 227
QY	359	N--QASKVHSEGLNEEDVISRODHEQTEKLELEIERCKEMIKAAQOQLLOQ 408
Db	228	DAQKGAVTSEATDRE-----LELE----AOLVEA-RSIIQEQ 260
RESULT 12		
AAG48365		
ID	AAG48365	standard; protein; 307 AA.
XX	AC	AAG48365;
XX	18-OCT-2000	(first entry)
DT	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 61067.
DE	XX	Protein identification; signal transduction pathway; metabolic pathway;
KW	XX	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	XX	termination sequence.
OS	XX	Arabidopsis thaliana.

XX EP1033405-A2.  
PN  
XX  
PD  
XX  
PF  
XX  
25-FEB-2000; 2000EP-00301439.  
25-FEB-1999; 99US-0121825P.  
05-MAR-1999; 99US-0123180P.  
09-MAR-1999; 99US-0123548P.  
23-MAR-1999; 99US-0125788P.  
25-MAR-1999; 99US-0126264P.  
29-MAR-1999; 99US-0126785P.  
01-APR-1999; 99US-0127462P.  
06-APR-1999; 99US-0128234P.  
08-APR-1999; 99US-0128714P.  
16-APR-1999; 99US-0128645P.  
19-APR-1999; 99US-0130077P.  
21-APR-1999; 99US-0130449P.  
23-APR-1999; 99US-0130510P.  
23-APR-1999; 99US-0130891P.  
28-APR-1999; 99US-0131449P.  
30-APR-1999; 99US-0132048P.  
30-APR-1999; 99US-0132407P.  
04-MAY-1999; 99US-0132484P.  
05-MAY-1999; 99US-0132485P.  
06-MAY-1999; 99US-0132486P.  
06-MAY-1999; 99US-0132487P.  
07-MAY-1999; 99US-0132863P.  
11-MAY-1999; 99US-0132566P.  
14-MAY-1999; 99US-0134218P.  
14-MAY-1999; 99US-0134219P.  
14-MAY-1999; 99US-0134221P.  
14-MAY-1999; 99US-0134370P.  
18-MAY-1999; 99US-0134768P.  
19-MAY-1999; 99US-0134941P.  
20-MAY-1999; 99US-0135124P.  
21-MAY-1999; 99US-0135353P.  
24-MAY-1999; 99US-0135629P.  
25-MAY-1999; 99US-0136021P.  
27-MAY-1999; 99US-0136392P.  
28-MAY-1999; 99US-0136782P.  
01-JUN-1999; 99US-0137222P.  
03-JUN-1999; 99US-0137528P.  
04-JUN-1999; 99US-0137502P.  
07-JUN-1999; 99US-0137724P.  
08-JUN-1999; 99US-0138094P.  
10-JUN-1999; 99US-0138540P.  
10-JUN-1999; 99US-0138847P.  
14-JUN-1999; 99US-0139119P.  
16-JUN-1999; 99US-0139452P.  
16-JUN-1999; 99US-0139453P.  
17-JUN-1999; 99US-0139492P.  
18-JUN-1999; 99US-0139454P.  
18-JUN-1999; 99US-0139455P.  
18-JUN-1999; 99US-0139456P.  
18-JUN-1999; 99US-0139457P.  
18-JUN-1999; 99US-0139458P.  
18-JUN-1999; 99US-0139459P.  
18-JUN-1999; 99US-0139460P.  
18-JUN-1999; 99US-0139461P.  
18-JUN-1999; 99US-0139462P.  
18-JUN-1999; 99US-0139463P.  
18-JUN-1999; 99US-0139750P.  
18-JUN-1999; 99US-0139763P.  
21-JUN-1999; 99US-0139817P.  
22-JUN-1999; 99US-0139899P.  
23-JUN-1999; 99US-0140353P.  
23-JUN-1999; 99US-0140354P.  
24-JUN-1999; 99US-0140695P.  
28-JUN-1999; 99US-0140823P.  
29-JUN-1999; 99US-0140991P.  
30-JUN-1999; 99US-0141287P.  
01-JUL-1999; 99US-0141842P.  
01-JUL-1999; 99US-0142154P.  
02-JUL-1999; 99US-0142055P.  
06-JUL-1999; 99US-0142390P.  
08-JUL-1999; 99US-0142803P.  
09-JUL-1999; 99US-0142920P.  
12-JUL-1999; 99US-0142977P.  
13-JUL-1999; 99US-0143542P.  
14-JUL-1999; 99US-0143624P.  
15-JUL-1999; 99US-0144005P.  
16-JUL-1999; 99US-0144085P.  
16-JUL-1999; 99US-0144086P.  
19-JUL-1999; 99US-0144325P.  
19-JUL-1999; 99US-0144331P.  
19-JUL-1999; 99US-0144332P.  
19-JUL-1999; 99US-0144333P.  
19-JUL-1999; 99US-0144334P.  
19-JUL-1999; 99US-0144335P.  
20-JUL-1999; 99US-0144352P.  
20-JUL-1999; 99US-0144632P.  
20-JUL-1999; 99US-0144884P.  
21-JUL-1999; 99US-0144814P.  
21-JUL-1999; 99US-0145086P.  
21-JUL-1999; 99US-0145088P.  
22-JUL-1999; 99US-0145085P.  
22-JUL-1999; 99US-0145087P.  
22-JUL-1999; 99US-0145089P.  
22-JUL-1999; 99US-0145132P.  
23-JUL-1999; 99US-0145145P.  
23-JUL-1999; 99US-0145218P.  
23-JUL-1999; 99US-0145224P.  
26-JUL-1999; 99US-0145276P.  
27-JUL-1999; 99US-0145913P.  
27-JUL-1999; 99US-0145918P.  
27-JUL-1999; 99US-0145919P.  
28-JUL-1999; 99US-0145951P.  
02-AUG-1999; 99US-0146386P.  
02-AUG-1999; 99US-0146388P.  
02-AUG-1999; 99US-0146389P.  
03-AUG-1999; 99US-0147038P.  
04-AUG-1999; 99US-0147204P.  
04-AUG-1999; 99US-0147302P.  
05-AUG-1999; 99US-0147192P.  
05-AUG-1999; 99US-0147260P.  
06-AUG-1999; 99US-0147303P.  
06-AUG-1999; 99US-0147416P.  
09-AUG-1999; 99US-0147453P.  
09-AUG-1999; 99US-0147935P.  
10-AUG-1999; 99US-0148171P.  
11-AUG-1999; 99US-0148319P.  
12-AUG-1999; 99US-0148341P.  
13-AUG-1999; 99US-0148565P.  
13-AUG-1999; 99US-0148684P.  
16-AUG-1999; 99US-0149368P.  
17-AUG-1999; 99US-0149175P.  
18-AUG-1999; 99US-0149426P.  
20-AUG-1999; 99US-0149722P.  
20-AUG-1999; 99US-0149723P.  
20-AUG-1999; 99US-0149929P.  
23-AUG-1999; 99US-0149902P.  
23-AUG-1999; 99US-0149930P.  
25-AUG-1999; 99US-0150566P.  
26-AUG-1999; 99US-0150884P.  
27-AUG-1999; 99US-0151065P.  
27-AUG-1999; 99US-0151086P.  
27-AUG-1999; 99US-0151080P.  
30-AUG-1999; 99US-0151303P.  
31-AUG-1999; 99US-0151438P.  
01-SEP-1999; 99US-0151930P.  
07-SEP-1999; 99US-0152363P.  
10-SEP-1999; 99US-0153070P.  
13-SEP-1999; 99US-0153758P.  
15-SEP-1999; 99US-0154018P.





PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-01499930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139750P.	PR	27-AUG-1999;	99US-0151065P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151066P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151080P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140353P.	PR	31-AUG-1999;	99US-0151438P.
PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	23-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0152363P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0153070P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153758P.
PR	30-JUN-1999;	99US-0141287P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0141842P.	PR	16-SEP-1999;	99US-0154039P.
PR	01-JUL-1999;	99US-0142154P.	PR	20-SEP-1999;	99US-0154779P.
PR	02-JUL-1999;	99US-0142055P.	PR	22-SEP-1999;	99US-0155139P.
PR	06-JUL-1999;	99US-0142390P.	PR	23-SEP-1999;	99US-0155486P.
PR	08-JUL-1999;	99US-0142803P.	PR	24-SEP-1999;	99US-0155659P.
PR	09-JUL-1999;	99US-0142920P.	PR	28-SEP-1999;	99US-0156458P.
PR	12-JUL-1999;	99US-0142977P.	PR	29-SEP-1999;	99US-0156596P.
PR	13-JUL-1999;	99US-0143542P.	PR	04-OCT-1999;	99US-0157117P.
PR	14-JUL-1999;	99US-0143624P.	PR	05-OCT-1999;	99US-0157753P.
PR	15-JUL-1999;	99US-0144005P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144086P.	PR	07-OCT-1999;	99US-0158029P.
PR	16-JUL-1999;	99US-0144086P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144325P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144331P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144334P.	PR	14-OCT-1999;	99US-0159329P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159637P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0144814P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145086P.	PR	21-OCT-1999;	99US-0160741P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160814P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145145P.	PR	22-OCT-1999;	99US-0160980P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160981P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160989P.
PR	26-JUL-1999;	99US-0145276P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161406P.
PR	27-JUL-1999;	99US-0145919P.	PR	26-OCT-1999;	99US-0161359P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146388P.	PR	28-OCT-1999;	99US-0161920P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161992P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147204P.	PR	29-OCT-1999;	99US-0162142P.
PR	04-AUG-1999;	99US-0147302P.			
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			
PR	09-AUG-1999;	99US-0147493P.			
PR	09-AUG-1999;	99US-0147935P.			
PR	10-AUG-1999;	99US-0148171P.			
PR	11-AUG-1999;	99US-0148319P.			
PR	12-AUG-1999;	99US-0148341P.			
PR	13-AUG-1999;	99US-0148565P.			
PR	13-AUG-1999;	99US-0148684P.			
PR	16-AUG-1999;	99US-0149368P.			
PR	17-AUG-1999;	99US-0149175P.			
PR	18-AUG-1999;	99US-0149426P.			
PR	20-AUG-1999;	99US-0149722P.			
PR	20-AUG-1999;	99US-0149723P.			
PR	23-AUG-1999;	99US-0149902P.			
			Query Match 8.4%; Score 265.5; DB 3; Length 288;		
			Best Local Similarity 25.7%; Pred. No. 2.3e-10;		
			Matches 75; Conservative 80; Mismatches 92; Indels 45; Gaps 12;		
QY	135	QSCYAKLKEQLETSRREMIGLQERDQL-----QCKN-----RSLHQLLKNKKEVQKLQ 184			
Db	7	QSDIARLEAKVE--RLEAL-LQCKDREIATITRTEAKNTAALKSQIEKLQERDEFORM- 62			
QY	185	NIASR--ATQYNHDVKKEREYNKLERLHOLVM-NKKDKNIAMDVLNYYVRADGKRG 241			
Db	63	-VIGNQVQAQOITHMKKKKXDYIKLQERLKNQVLMEKKESRSGMEIWNLLQKERQRT 121			
QY	242	WRTDKTEARNEDEMYKILLNDYERQKQILMENAEKVKLQOMKKEMISLLSPQKKPRE 301			
Db	122	WNGKKTDT----DFYKIVDAYEAKNQELMAENTSRLALLSMQTDMDRDLN----- 169			
QY	302	RAEDGTGTVAJSDIEDDSGELS-----RDSVNGLSCDTTRQELTNSIRKQWRLLKSHVKL 357			

Db 170 -APNGSATLAGEKHEADPSQSPGKGTDFDLPYRMARQIEESLRTKMASIKESMVQL 228  
QY 358 DNOASKVHSEGLNEEDVIRQDHEQTEKLELEIERCKEMIKAQOOLQOOL 409  
Db 229 QDAPKRAS-----VTSEATERELE-LEAQLVEARSIIOREQESIMSKHL 270

RESULT 14  
AAG33166  
ID AAG33166 standard; protein; 277 AA.  
XX AC AAG33166;  
XX AC  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 40145.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.



PR 01-JUN-1999;	99US-0137322P.	PR 05-AUG-1999;	99US-0147192P.
PR 03-JUN-1999;	99US-0137528P.	PR 05-AUG-1999;	99US-0147260P.
PR 04-JUN-1999;	99US-0137502P.	PR 06-AUG-1999;	99US-0147303P.
PR 07-JUN-1999;	99US-0137724P.	PR 06-AUG-1999;	99US-0147416P.
PR 08-JUN-1999;	99US-0138094P.	PR 09-AUG-1999;	99US-0147493P.
PR 10-JUN-1999;	99US-0138540P.	PR 09-AUG-1999;	99US-0147935P.
PR 10-JUN-1999;	99US-0138847P.	PR 10-AUG-1999;	99US-0148171P.
PR 14-JUN-1999;	99US-0139119P.	PR 11-AUG-1999;	99US-0148319P.
PR 16-JUN-1999;	99US-0139452P.	PR 12-AUG-1999;	99US-0148341P.
PR 17-JUN-1999;	99US-0139453P.	PR 13-AUG-1999;	99US-0148565P.
PR 17-JUN-1999;	99US-0139492P.	PR 13-AUG-1999;	99US-0148565P.
PR 18-JUN-1999;	99US-0139454P.	PR 16-AUG-1999;	99US-0148684P.
PR 18-JUN-1999;	99US-0139455P.	PR 16-AUG-1999;	99US-0149368P.
PR 18-JUN-1999;	99US-0139456P.	PR 17-AUG-1999;	99US-0149175P.
PR 18-JUN-1999;	99US-0139457P.	PR 18-AUG-1999;	99US-0149426P.
PR 18-JUN-1999;	99US-0139458P.	PR 20-AUG-1999;	99US-0149722P.
PR 18-JUN-1999;	99US-0139459P.	PR 20-AUG-1999;	99US-0149723P.
PR 18-JUN-1999;	99US-0139460P.	PR 20-AUG-1999;	99US-0149929P.
PR 18-JUN-1999;	99US-0139461P.	PR 23-AUG-1999;	99US-0149902P.
PR 18-JUN-1999;	99US-0139462P.	PR 23-AUG-1999;	99US-0149930P.
PR 18-JUN-1999;	99US-0139463P.	PR 25-AUG-1999;	99US-0150566P.
PR 18-JUN-1999;	99US-0139750P.	PR 26-AUG-1999;	99US-0150884P.
PR 18-JUN-1999;	99US-0139763P.	PR 27-AUG-1999;	99US-0151065P.
PR 21-JUN-1999;	99US-0139817P.	PR 27-AUG-1999;	99US-0151066P.
PR 22-JUN-1999;	99US-0139899P.	PR 27-AUG-1999;	99US-0151080P.
PR 23-JUN-1999;	99US-0140353P.	PR 30-AUG-1999;	99US-0151303P.
PR 23-JUN-1999;	99US-0140354P.	PR 31-AUG-1999;	99US-0151438P.
PR 24-JUN-1999;	99US-0140695P.	PR 01-SEP-1999;	99US-0151930P.
PR 28-JUN-1999;	99US-0140823P.	PR 07-SEP-1999;	99US-0152363P.
PR 29-JUN-1999;	99US-0140991P.	PR 10-SEP-1999;	99US-0153070P.
PR 30-JUN-1999;	99US-0141287P.	PR 13-SEP-1999;	99US-0153758P.
PR 01-JUL-1999;	99US-0141842P.	PR 15-SEP-1999;	99US-0154018P.
PR 01-JUL-1999;	99US-0142154P.	PR 16-SEP-1999;	99US-0154039P.
PR 02-JUL-1999;	99US-0142055P.	PR 20-SEP-1999;	99US-0154779P.
PR 06-JUL-1999;	99US-0142390P.	PR 22-SEP-1999;	99US-0155139P.
PR 08-JUL-1999;	99US-0142803P.	PR 23-SEP-1999;	99US-0155486P.
PR 09-JUL-1999;	99US-0142920P.	PR 24-SEP-1999;	99US-0155659P.
PR 12-JUL-1999;	99US-0142977P.	PR 28-SEP-1999;	99US-0156458P.
PR 13-JUL-1999;	99US-0143542P.	PR 29-SEP-1999;	99US-0156596P.
PR 14-JUL-1999;	99US-0143624P.	PR 04-OCT-1999;	99US-0157117P.
PR 15-JUL-1999;	99US-0144005P.	PR 05-OCT-1999;	99US-0157753P.
PR 16-JUL-1999;	99US-0144085P.	PR 06-OCT-1999;	99US-0157865P.
PR 16-JUL-1999;	99US-0144086P.	PR 07-OCT-1999;	99US-0158029P.
PR 19-JUL-1999;	99US-0144325P.	PR 08-OCT-1999;	99US-0158232P.
PR 19-JUL-1999;	99US-0144331P.	PR 12-OCT-1999;	99US-0158369P.
PR 19-JUL-1999;	99US-0144332P.	PR 13-OCT-1999;	99US-0159293P.
PR 19-JUL-1999;	99US-0144333P.	PR 13-OCT-1999;	99US-0159294P.
PR 19-JUL-1999;	99US-0144334P.	PR 14-OCT-1999;	99US-0159295P.
PR 20-JUL-1999;	99US-0144335P.	PR 14-OCT-1999;	99US-0159329P.
PR 20-JUL-1999;	99US-0144352P.	PR 14-OCT-1999;	99US-0159330P.
PR 20-JUL-1999;	99US-0144632P.	PR 14-OCT-1999;	99US-0159331P.
PR 21-JUL-1999;	99US-0144884P.	PR 14-OCT-1999;	99US-0159637P.
PR 21-JUL-1999;	99US-0144814P.	PR 18-OCT-1999;	99US-0159638P.
PR 21-JUL-1999;	99US-0145086P.	PR 21-OCT-1999;	99US-0159584P.
PR 21-JUL-1999;	99US-0145088P.	PR 21-OCT-1999;	99US-0160741P.
PR 22-JUL-1999;	99US-0145085P.	PR 21-OCT-1999;	99US-0160767P.
PR 22-JUL-1999;	99US-0145087P.	PR 21-OCT-1999;	99US-0160768P.
PR 22-JUL-1999;	99US-0145089P.	PR 21-OCT-1999;	99US-0160770P.
PR 22-JUL-1999;	99US-0145192P.	PR 21-OCT-1999;	99US-0160814P.
PR 23-JUL-1999;	99US-0145145P.	PR 22-OCT-1999;	99US-0160815P.
PR 23-JUL-1999;	99US-0145218P.	PR 22-OCT-1999;	99US-0160980P.
PR 23-JUL-1999;	99US-0145224P.	PR 22-OCT-1999;	99US-0160981P.
PR 26-JUL-1999;	99US-0145276P.	PR 25-OCT-1999;	99US-0160989P.
PR 27-JUL-1999;	99US-0145913P.	PR 25-OCT-1999;	99US-0161404P.
PR 27-JUL-1999;	99US-0145918P.	PR 25-OCT-1999;	99US-0161405P.
PR 27-JUL-1999;	99US-0145919P.	PR 25-OCT-1999;	99US-0161406P.
PR 28-JUL-1999;	99US-0145951P.	PR 26-OCT-1999;	99US-0161359P.
PR 02-AUG-1999;	99US-0146386P.	PR 26-OCT-1999;	99US-0161360P.
PR 02-AUG-1999;	99US-0146388P.	PR 26-OCT-1999;	99US-0161361P.
PR 02-AUG-1999;	99US-0146389P.	PR 28-OCT-1999;	99US-0161920P.
PR 03-AUG-1999;	99US-0147038P.	PR 28-OCT-1999;	99US-0161992P.
PR 04-AUG-1999;	99US-0147204P.	PR 28-OCT-1999;	99US-0161993P.
PR 04-AUG-1999;	99US-0147302P.	PR 29-OCT-1999;	99US-0162142P.



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 13, 2006, 23:13:44 ; Search time 8205 Seconds  
(without alignments)  
7189.701 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVDPVLCTENKLS.....CYSGSSAFSAHGRRDRLP 615

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+\_p2n.model -DEV=xlp  
-Q=/abss/ABSSWEB\_spool/US10644084/runat\_12062006\_150414\_1351/app\_query.fasta\_1  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03p  
-USER=US10644084@cgn\_1\_8328 @runat\_12062006\_150414\_1351 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_env.\*  
2: gb\_pat.\*  
3: gb\_ph.\*  
4: gb\_pl.\*  
5: gb\_pr.\*  
6: gb\_ro.\*  
7: gb\_sts.\*  
8: gb\_sy.\*  
9: gb\_un.\*  
10: gb\_vi.\*  
11: gb\_ov.\*  
12: gb\_atg.\*  
13: gb\_in.\*  
14: gb\_om.\*  
15: gb\_ba.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	3165	100.0	2692	6 AF532969	
2	3165	100.0	3424	6 AK173064	
3	3165	100.0	3425	6 BC021749	

#### SUMMARIES

4	3162	99.9	3410	6 BC031527	BC031527 Mus muscu
5	2937	92.8	2547	6 BC078687	BC078687 Rattus no
6	2937	92.8	3195	6 AF532970	AF532970 Rattus no
7	2786.5	88.0	3220	5 BC033637	BC033637 Homo sapi
8	2786.5	88.0	5835	5 C0722034	C0722034 Sequence
9	2786.5	88.0	5835	5 AB023140	AB023140 Homo sapi
10	2781.5	87.9	2379	5 AY367055	AY367055 Homo sapi
11	2720	85.9	2272	2 AR718997	AR718997 Sequence
12	2720	85.9	2272	2 AX747972	AX747972 Sequence
13	2720	85.9	2272	5 AK093173	AK093173 Homo sapi
14	2691	85.0	2540	5 BC064389	BC064389 Homo sapi
15	2514	79.4	3073	5 AY651262	AY651262 Homo sapi
16	1982.5	62.6	2716	2 BD156617	BD156617 Primer fo
17	1982.5	62.6	2716	2 AX877359	AX877359 Sequence
18	1982.5	62.6	2716	5 AK001710	AK001710 Homo sapi
C 19	1754	55.4	185929	12 AC017085	AC017085 Homo sapi
C 20	1754	55.4	188840	5 AC046144	AC046144 Homo sapi
C 21	1558.5	49.2	3218	11 BC070587	BC070587 Xenopus l
C 22	1554	49.1	3892	11 BC070749	BC070749 Xenopus l
C 23	1371.5	43.3	2594	11 BC060934	BC060934 Danio rer
C 24	1192	37.7	826	2 BD148918	BD148918 Primer fo
C 25	1192	37.7	826	2 AX868856	AX868856 Sequence
C 26	1037.5	32.8	781	11 BX934353	BX934353 Gallus ga
C 27	856	27.0	1503	2 BD158988	BD158988 Primer fo
C 28	856	27.0	1503	2 AX881380	AX881380 Sequence
C 29	856	27.0	1503	5 AK021515	AK021515 Homo sapi
C 30	777	24.5	258641	12 AC127120	AC127120 Rattus no
C 31	728.5	23.0	674	11 CR353768	CR353768 Gallus ga
C 32	715	22.6	563	2 BD150903	BD150903 Primer fo
C 33	715	22.6	563	2 AX870841	AX870841 Sequence
C 34	671.5	21.2	498	2 BD058793	BD058793 Secreted
C 35	650	20.5	403	2 AX071986	AX071986 Sequence
C 36	642.5	20.3	1107	11 BC083288	BC083288 Danio rer
C 37	635.5	20.1	428	2 BD026751	BD026751 Sequence
C 38	635.5	20.1	428	2 AR727012	AR727012 Sequence
C 39	635.5	20.1	428	2 AX887141	AX887141 Sequence
C 40	627.5	19.8	227133	6 AL663098	AL663098 Mouse DNA
C 41	530	16.7	165330	6 AC124987	AC124987 Mus muscu
C 42	520.5	16.4	233457	12 AC106173	AC106173 Rattus no
C 43	520.5	16.4	255776	12 AC142185	AC142185 Rattus no
C 44	520.5	16.4	306870	12 AC098557	AC098557 Rattus no
C 45	438	13.8	161874	5 AC114482	AC114482 Homo sapi

#### ALIGNMENTS

RESULT 1	AF532969	AF532969	2692 bp	mRNA	linear	ROD 03-FEB-2003
LOCUS	AF532969	Mus musculus afadin- and alpha-actinin-binding protein ADIP mRNA,				
DEFINITION	AF532969	complete cds.				
ACCESSION	AF532969	AF532969.1	GI:27451838			
VERSION	AF532969					
KEYWORDS	AF532969	Mus musculus (house mouse)				
SOURCE	AF532969	Mus musculus				
ORGANISM	AF532969	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
REFERENCE	AF532969	1 (bases 1 to 2692)				
AUTHORS	AF532969	Asada,M., Irie,K., Morimoto,K., Yamada,A., Ikeda,W., Takeuchi,M. and Takai,Y.				
TITLE	AF532969	ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized at Cell-Cell Adherens Junctions				
JOURNAL	AF532969	J. Biol. Chem. 278 (6), 4103-4111 (2003)				
PUBMED	AF532969	12446711				
REFERENCE	AF532969	2 (bases 1 to 2692)				
AUTHORS	AF532969	Asada,M., Irie,K. and Takai,Y.				
TITLE	AF532969	Direct Submission				
JOURNAL	AF532969	Submitted (26-JUL-2002) Department of Molecular Biology and Biochemistry, Osaka University Graduate School of Medicine/Faculty of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan				
FEATURES	AF532969	Location/Qualifiers				

source	1. 2692 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6" /db_xref="taxon:10090" 80..1927 /codon_start=1 /product="afadin- and alpha-actinin-binding protein ADIP" /protein_id="AAO15015.1" /db_xref="GI:27451839" translation="MGDMVTVDPLCTENKNLSQYTSKMPSSILYSQQVLCSSVP LSKNVGVFGVCFGENIEQISYLDQELTTFGPPSYLESKKEAKRELNIYAVLNC MVELLVLQRNLAAQESVETQNLKLGSDMDHLOSCYAKLKEOLETSREMIGLQERDR QLQCKNSLHQLLNKEDVEQKLQNIIASRATQYNHVDVKKEREYNKLERLHQLVMN LKQNIANDVINYGRADRGKRSWRDTEARNEDMYKILLNDIETFRQKQILMENAE LKKYLQMKEMISLLSPQKKPREAEDGTGTVAISDIEDSGELSRVWGLSCDT VREQTLSIRKQWILKSHVEKLDNQASKHSEGLNEEDVISRDHQEETKEKLEITE RCKMIKAQOOLLQOOLATTCDDDTTSLLRDCLLEBEKRLKEWTLFKEQKNFPER RRSFTEAAIRGLERKAFERERASWVKQOFLNTPDNPHQNSNVKLPFAFGSGSDPDN LIVHSRPRQKLHSHVANGVPACTSKLTKSLPASPSSTDFRQTHSCVSEHSISVINIT PEESKPEVARESTDQRKWSVQSRPSREGCTSGCASSAFRSAHGRDDLP"
CDS	
ORIGIN	
Alignment Scores:	
Pred. No.:	0
Score:	3165.00
Percent Similarity:	100.0%
Best Local Similarity:	100.0%
Query Match:	100.0%
DB:	6
US-10-644-084-2 (1-615) x AF532969 (1-2692)	
Qy	1 MetGlyAspThrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db	80 ATGGGAGATTGGATGACTGTACAGATCCAGTCTGTGTACAGAAAACAAAACTCTCT 139
Qy	21 GlnTyrThrSerGluThrLysMetSerProSerSerLysThrSerGlnValLeuCys 40
Db	140 CAATATACCTCAGAAACAAAGATGTCTCCGTCCTAGTTGTACTCCCAAGCAAGTTCTGTGC 199
Qy	41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db	200 TCTTCAGTACCTTTATCCAAAAACGTGATGGTGTCTTCGGTGTCTTCGACAGGAGAG 259
Qy	61 AsnIleGluGlnSerIleSerTyLysAspGlnGluLeuThrThrPheGlyPheProSer 80
Db	260 AACATTGAACAAAGTATTCTCTATCTTGTATCAGGAGCTGACCACTTTCGGGTTTCTCTCC 319
Qy	81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db	320 TTGTATGAAGAATCCAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTCTGT 379
Qy	101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db	380 AACTGTATGAACGAGCTGCTGCTCTTCAGCGGAAGAACCTGCTGGCCAGGAGAGCGTG 439
Qy	121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLys 140
Db	440 GAGACACAGAACCTTGAAGCTGGGAGTGCATGACATGGACCACTTGCAGAGCTGTACGCCAAA 499
Qy	141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db	500 CTTAAGGAGCAGTTGGAAACGTCCAGCGGGGAGATGATCGGGCTTCAAGAGAGAGACAGG 559
Qy	161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db	560 CAGCTGCAGTCCAGAACACAGAGATTGCATCAGCTCCTGAGAGATGAGAAAGATGAGGTA 619
Qy	181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db	620 CAAAAATACAAAAATATCATAGCCAGCGCGGTACTCAGTATATATCATGATGTGAAGAGG 679
Qy	201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db	680 AAGAGCGGTGAATATATTAAGCTTAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAG 739
Qy	221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db	740 GATAAAAAACATAGCCATGGATGTTTAAATATATGTGGTCCAGCTGATGGCAACAGGAGC 799
Qy	241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db	800 TCATGGAGGACTGACAAAACAGAACCCAGGAATGAAGATGAGATGTACAAAATTTCTGTG 859
Qy	261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db	860 AATGATTATGAGTACCCCAAGAGACAGATCTCTGATGGAGAACCGGAGCTGAAGAAGGTC 919
Qy	281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db	920 CTCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCTCAGAAAGAAAGACCCAGG 979
Qy	301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
Db	980 GAAAGAGCAGAGGACGACAGGACTGTTGCTATCTCCGATATAGAAGATGACTCTGGG 1039
Qy	321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db	1040 GAACTGAGCAGACAGACAGCTGTGGGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1099
Qy	341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db	1100 AACAGCATCAGGAAACAGTGGAGAAATTTGAAAGTCATGTAGAAAACTCGATAACCAA 1159
Qy	361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db	1160 GCTTCGAAGGTACACTCAGAGGGCTTAAATGAGAGAGACGTCACTCTCAGCAACAAGACAT 1219
Qy	381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Db	1220 GACACAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1279
Qy	401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db	1280 CAGCAGCAGCTCTTACAGCAGCAGCTGCCCAACCCAGCTGTGATGATGACACCACTCACTG 1339
Qy	421 LeuArgAspCysTyrLeuLeuLeuGluLysGluArgLeuLysGluTyrThrLeuPhe 440
Db	1340 TTGCGAGACTGTACTTGTCTGGAGAAAGAGAACGCCCTTAAAGAAAGAGTGACCCCTTTT 1399
Qy	441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db	1400 AAAGACAAAAAAGAAATTTTGAGAGAGAAAGCGAAGCTTTACAGAAGCTGCCATTCGA 1459
Qy	461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db	1460 TTGGGTTGGAGAGAAAGCGCTTTGAAGAGAGAGCGAGCTGGGTGAAGACAGCAGCTTT 1519
Qy	481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db	1520 TTAACAATGACGAACCTTTTGACCAACCACTCAGAAATGTGAAACCTTTTCAGTGCCTTC 1579
Qy	501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db	1580 TCAGGAAGTTCTGATCCAGACAACTTATATAGTCCACTCACGGCCACGGCAAAAGAGCTA 1639
Qy	521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db	1640 CACAGTGTGGCTAATGGGGTGCAGCTTGACACTCAAAAACCTGACTAAATCTCTCTCCGCC 1699
Qy	541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db	1700 TCACCTTCTACTCAGACTTTCGCCAGACACATTCATGTGTGTGTGAACACAGTTCATC 1759
Qy	561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaIleArgGluSerThr 580





Qy	301	GluArgAlaGluAaspGlyThrGlyThrValAlaIleSerAspIleGluAaspSerGly	320
Ds	1172	GAAGAGCAGAGGACGGCAGCAGGACCTGTTGCTATCTCCGATATAGAGATGACTCTGGG	1231
Qy	321	GluLeuSerArgAaspSerValTrrpGlyLeuSerCysAaspThrValArgGluInLeuThr	340
Ds	1232	GAATGAGCAGACAGCGGTGGGGCTTTCCTGTACACTGTGAGAGAGCAGCTGACA	1291
Qy	341	AsnSerIleArgLysGlnTrrpArgIleLeuLysSerHisValGluLysLeuAaspAsnGln	360
Ds	1292	AACAGCATCAGAAACAGTGGAGAAATTTGAAAGTCATGTAGAAAAAATCGATAACAA	1351
Qy	361	AlaSerLysValHisSerGluLysGluLeuAsnGluAaspValIleSerArgGlnAspHis	380
Ds	1352	GCTTCGAAGTACACTCAGAGGGCTTAATAGAGAGGACGTCATCTCAGCAACAGCAT	1411
Qy	381	GluGlnGluThrGluLysLeuGluLysGluIleGluArgCysLysGluMetIleLysAla	400
Ds	1412	GAGCAGAGACTGAGAAACTGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGCT	1471
Qy	401	GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAaspAaspThrThrSerLeu	420
Ds	1472	CAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACTGTGTATGATGACACCACTCACTG	1531
Qy	421	LeuArgAaspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTrrpThrLeuPhe	440
Ds	1532	TTGGGAGACTGTTACTTGTCTGGAAGAAAGACGCTTAAAGAGAGTGGACCTTTTT	1591
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAilaileArg	460
Ds	1592	AAAGAGCNAANAAGAAATTTTGAGAGAGAAAGGCGAAGCTTTACAGAACTGCCATTCGA	1651
Qy	461	LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrrpValLysGlnGlnPhe	480
Ds	1652	TTGGGGTTGGAGAGAAAGCGCTTTGAGAGAGCAGCAGCTGGTAAAGCAGCAGTTT	1711
Qy	481	LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe	500
Ds	1712	TTAAACATGACGAACTTTGACCAACAGAACTCAGAAATGTGAAACTTTTCAGTGCCTTC	1771
Qy	501	SerGlySerSerAaspProAaspAsnLeuIleValHisSerArgProArgGlnLysLysLeu	520
Ds	1772	TCAGGAAGTCTGATCCAGACAACTTATATAGTCCACTCAGCGCCACGGCAAGAGCTA	1831
Qy	521	HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540
Ds	1832	CACAGTGTGGCTAATGGGTGCCAGCTTGCCACATCAAAACTGACTAAATCTCTTCCTGCC	1891
Qy	541	SerProSerThrSerAaspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560
Ds	1892	TCACCTTCTACTCAGACTTTTCGCCAGACACATTCATGTGTCTGAAACAGTTCATC	1951
Qy	561	SerValLeuAsnIleThrProGluSerLysProSerGluValAlaArgGluSerThr	580
Ds	1952	AGTGTGCTGAATATAACTCTTGAGAAAGTAAACCACTGAGGTGCAAGAGAAAGCAGC	2011
Qy	581	AspGlnLysTrrpSerValGlnSerArgProSerArgGluGlyCysTyrSerGlyCys	600
Ds	2012	GATCAGAAAGTGGAGCGTGCAGTTCGAGGCCAGCTCGCGGAGGGGTGTACAGCGGATGC	2071
Qy	601	SerSerAlaPheArgSerAlaHisGlyAspArgAaspLeuPro	615
Ds	2072	TCCTCGGCTTTTCAGGAGCGCTCAGGGGACCGAGATGATGATCTTACCT	2116
RESULT 3			
BC021749			
LOCUS	BC021749	3425 bp	mRNA linear ROD 18-JUL-2005
DEFINITION	Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,		
	mRNA (cdna clone MGC:25823 IMAGE:4165430), complete cds.		
ACCESSION	BC021749		
VERSION	BC021749.1	GI:18256805	
KEYWORDS	MGC.		

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Euarchotheria; Glirres; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus. 1 (bases 1 to 3425)
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klauener, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Platchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, E., Soates, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Udwin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.O., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Young, A.C., Shevchenko, Y., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A. Mammalian Gene Collection Program Team
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26): 16999-16902 (2002)
PUBMED	12575822
REFERENCE	
AUTHORS	
CONSRM	
TITLE	
JOURNAL	
REMARK	
COMMENT	

NTH MGC Project  
Direct Submission  
Submitted (18-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NTH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-@email.nih.gov](mailto:cgapbs-@email.nih.gov)  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sheed, A.J., Martin, R.G., Muzny, D.M., Nahanati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 30 Row: m Column: 16.  
Location/Qualifiers  
1. .3425  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:25823 IMAGE:4165430"  
/tissue\_type="Salivary gland, 10 week old female mouse"  
/clone\_lib="NCI CGAP\_SG2"  
/lab\_host="DH10B"  
/note="vector: pCMV-SPORT6"  
1. .3425  
/gene="Ssx2ip"  
/note="synonym: Adip"  
/db\_xref="GeneID:99167"  
/db\_xref="WGI:2139150"  
240..2087  
/gene="Ssx2ip"  
/codon\_start=1  
/product="synovial sarcoma, X breakpoint 2 interacting

protein"  
/db\_xref="GI:18256806"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
/translation="MGDMWTVDPVLCTENKNLSOYTSSTKSPSSLSYQQVLCSSVP  
LSKNVHGVFGVCTCGENLEQSI SYLDQBLTTFGFSFLYEESKEAKRELNIYAVINC  
WNEELVLQRKNLLAQESVETQNLKLGSDWDHLLQSCYAKLKEQLETSRREMIGLOERDR  
QIQCKNRSJLHJLKNKEQVEQKQNI IASRATQYNDHVRKEREYKNLKERLHQLVMN  
KKDKNMDVLNVRADGKGRSMDTKTEARNEDEMYKILLNDYERQKOTLMENAE  
LKKVLQMKKEMISLSQKKPRERAEDGTCTVAISDIEDDSGELSRSVWGLSCDT  
VREQLTNRKQWRILKSHVEKLNQAKVHSEGLNBDVLSRQDHEOETKLELEIE  
RCKEMIAQQQLQQLATTCDDDTTSLLRDCYLLKEERLKEEWTLFKEQKXPFERE  
RRSFTEAIRLGLERKAPEERASWVKQFLNMNFDQNSNVKLFSAFGSSDPDN  
LIVHRKQKQLHSHVANGVFACTSKLTSLPASPSSTSDFRQTHSCVSHSHSISVLNIT  
PEESKPEVARESTDQRKSVQSRPSREGCYGSCSSAFSAHGRDRDLP"

## ORIGIN

## Alignment Scores:

Pred. No.: 0 Length: 3425  
Score: 3165.00 Matches: 615  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 100.0% Indels: 0  
DB: 6 Gaps: 0

US-10-644-084-2 (1-615) x BC021749 (1-3425)

QY	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
DB	240	ATGGGAGATTGGATGACTGTGACACATCCAGTTCTGTGTACGAAACACAAAAATCTCTCT	299
QY	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
DB	300	CAATATACCTCAGAAACAAAGATGCTCCGTCAGTTTGTACTCCAGCAAGTTCTGTGC	359
QY	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
DB	360	TCCTTCAGTACCTTTATCCAAAACGTGATGTTGTTTCGGTGTCTCTGACACAGGAGAG	419
QY	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
DB	420	AACATTGAACAAAGTATTTCTTATCTTGATCAGGAGCTGACCACCTTCGGGTTCCTTCC	479
QY	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu	100
DB	480	TTGTATGAAGAATCCAAAGTAAAGAGCAAGAGAGAATTAATATATAGTCGCTGCTCG	539
QY	101	AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal	120
DB	540	AACTGTATGAACGAGCTGCTCGTGTCTCAGCGGAAGAACTTGCTGGCCACGAGAGCGTG	599
QY	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnInSerCysTyrAlaLys	140
DB	600	GAGACACAGAACTTGAAGCTGGCNGTGCATGATGNCCACCTTGACAGAGCTGTACGCCANA	659
QY	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGlnArgAspArg	160
DB	660	CTTAAGGAGCAGTTTGAACACCTCCAGCGCGGAGATGATCGGGCTTCAAGAGAGACAGG	719
QY	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal	180
DB	720	CAGCTGCAGTGCAAGAACAGAGAGTTTGCATCAGCTCTCTGAAAGATGAGAAAGATGAGGTA	779
QY	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg	200
DB	780	CAAAAAATTACAAATATATATAGCCAGCCGGGCTACTCAGTATATATCATGATGTGAAGAGG	839
QY	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys	220
DB	840	AAGGAGCGTGAAATATAAAGCTAAAGAGGCGCTGCATCAGCTCGTTTATGAACAAGAG	899
QY	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240

DB	900	GATAAAAAACATAGCCATCGATGTTTAAATATATGTGGTTCGAGCTGATGGCAACGAGGC	959
QY	241	SerTrpAlaGlnThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
DB	960	TCATGGAGAGCTGACAAAACAGAACCCAGGAATGAAGATGAGATGTACAAAATCTCTGTG	1019
QY	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280
DB	1020	AATGATTATGAGTACCGCCAGAGCAGATCTCTGATGGAGAACCGCGAGCTGAAGAAGGTC	1079
QY	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
DB	1080	CTCCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1139
QY	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320
DB	1140	GAAAGAGCAGAGGACGCGCACAGGCACTGTTGCTATCTCCGATATAGAAGATGACTCTGGG	1199
QY	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
DB	1200	GAACTGAGCAGACAGCGCTGTGGGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA	1259
QY	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360
DB	1260	AACAGCATTCAGGAACACAGTGGAGAAATTTTGAAGATGTCATGTAGAAAACTCGATAACCAA	1319
QY	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis	380
DB	1320	GCTTCGAAGGTACACTCAGAGGGCTTATGAGGAGGACGTCATCTCAGCACAGACCAT	1379
QY	381	GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla	400
DB	1380	GAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGCT	1439
QY	401	GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu	420
DB	1440	CAGCAGCAGCTCTTTACAGCAGCAGCTGGCCACCGTGTGATGATGACACACCTCTCCTG	1499
QY	421	LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluTyrThrLeuPhe	440
DB	1500	TTGGCAGACTGTTTACTTGTGTAAGAAAGAAAGAAAGCCCTTAAAGAAAGAGTGACCCCTTTT	1559
QY	441	LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg	460
DB	1560	AAAGCAAAAGAAAGAAATTTTGAGAGAGAAAGGCGAAGCTTTTACAGAAAGCTGCCATT	1619
QY	461	LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe	480
DB	1620	TTGGGGTTGGAGAGAAAGGGCGTTTGAAGAGAGCGAGCCAGCTGGGTAAAGCAGCAGTTT	1679
QY	481	LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe	500
DB	1680	TTAAACATGACGAACCTTTGACCCACAGAACTCAGAAATGTGAACCTTTTTCAGTGTCTTTC	1739
QY	501	SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu	520
DB	1740	TCAGGAAGTTCTGATCCAGACAACTTTATAGTCCACTCACGCCACCGCAAGAAAGAGCTA	1799
QY	521	HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540
DB	1800	CACAGTGTGGCTAATATGGGGTGCACACTTGACAACTCAAAACCTGACTAAATCTCTCTCT	1859
QY	541	SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560
DB	1860	TCACCTTCTACTTACAGACTTTCGCCAGACACATTCATGTGTCTGAAACAGATTCATC	1919
QY	561	SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaIleArgGluSerThr	580
DB	1920	AGTGTGCTGAATATACTCTGAGAAAGTAAACCAAGTGAGGTTCGCAAGAGAAAGCAGC	1979
QY	581	AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys	600

Db 1980 GATCAGAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGGAGGGGTGCTACAGCGGATGC 2039  
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
|||||  
Db 2040 TCCTCGGCGCTTCAGGAGCGCTCACGGGAGCGGATGACTTACCT 2084

RESULT 4  
BC031527  
LOCUS BC031527 3410 bp mRNA linear ROD 29-JUN-2004  
DEFINITION Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,  
mRNA (cDNA clone MGC:28268 IMAGE:4008624), complete cds.  
ACCESSION BC031527  
VERSION BC031527.1 GI:21594536

KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 3410)

REFERENCE  
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalios, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 3410)

TITLE Strausberg, R.  
JOURNAL Direct Submission  
PUBMED Submitted (08-JUN-2002) National Institutes of Health, Mammalian  
REFERENCE Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

REMARK  
COMMENT

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 36 Row: k Column: 10  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 20270272.  
Location/Qualifiers  
1. .3410  
/organism="Mus musculus"  
/mol\_type="mRNA"

FEATURES  
source

/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="MGC:28268 IMAGE:4008624"  
/tissue\_type="Mammary tumor metastatized to lung.  
MMTV-LTR/Wnt1 model. Expression driven by an MMTV-LTR  
enhancer."  
/clone\_lib="NCI\_CGAP\_Lu30"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. .3410  
/gene="Sex2ip"  
/note="synonym: Adip"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
248..2095  
/gene="Sex2ip"  
/codon\_start=1  
/product="synovial sarcoma, X breakpoint 2 interacting  
protein"  
/protein\_id="AAH31527.1"  
/db\_xref="GI:21594537"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
/translation="MGDMVTVDVPLCTENKNLSQYTSFTKMPSSLYSQVLCSSVP  
LSRVHGVFGVCTGENIEQSISLDQELTTFGPSLYEESKKEAKRELNIIVAVLN  
MNEILLQRKNLLAQSVEONKLGSDMDHLOSCKYAKLEQLETSRRREMLGQERDR  
OLOCKNRSLLHLLKNEKDEKONKNTIISARTOYNHDVKREYNKLERLHOLVNN  
KDKXNTAMVUNYVGRADKRGSRWRTDKTEARNEDEMYKILLNDYERQKILMENAE  
LKKVLOOMKEMISLSPKPKPRERADGTGTVASIDIDDSGEISRVSVMGSCDT  
VREQLTNSRKQWRILSHVEDLNQAKSHVSEGLNEEDVISRDHQEETKLELELE  
RKEMTKAQOQLLQOQLATTCDDDTTSLRDCLYLLKEERLKEEWTLFKQKNKFERE  
RSFTEAARLGLERKAFFERASWVKQOFLNMNFNDHNSENVKILFSFSGSDPDN  
LIVHSRPROKLLSHVANGVPACTSKLTLPASPTSDPROTHSCVSEHSISVLNIT  
PEESKSEVARESTDQKMSVQSPSSREGCYSCSSAFRSRHAHGRDRLP"

ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 3410  
Score: 3162.00 Matches: 614  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 99.9% Indels: 0  
DB: 6 Gaps: 0  
US-10-644-084-2 (1-615) x BC031527 (1-3410)  
Qy 1 MetGlyAspTrpMetThrValThrAppProValLeuCysThrGluAnLysAsnLeuSer 20  
Db 248 ATGGGAGATTGGATGACTGTGCAGATCCAGTTCGTGTACAGAAACAAATCTCTCT 307  
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Db 308 CAATATACCTCAGAAACAAAGATGTCTCGCTCAGTTTGTACTCCCAAGCAAGTTCTGTGC 367  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 368 TCTTCAGTACCTTTATCCAGAAATGTGATGGTGTTCCTCGGTGTCTTCGCACAGGAGAG 427  
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
Db 428 AACATTGAACAAAGATATTCCTATCTTGATCAGAGCTGACCACTTCGGGTTCCTTCC 487  
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Db 488 TTGTATGAAGATCCAAAGATTAAGAGGCAAGAGAGAAATTAATATAGTCGTGTTCTG 547  
Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
Db 548 AACTGTATGAACAGAGTGTCTCGTGTTCAGCGGAAGAACCTGTGTCGCCACAGAGAGGTG 607  
Qy 121 GluThrClnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db 608 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGACCCTGCAGAGCTGCTACGCCAAA 667

QY 141 LeuLysGluGlnLeuGlnThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 DB CTTAAGGAGCAGTTGGAAACGTCACGCGGAGATGATCGGGCTTCAAGAGAGACAGG 727  
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180  
 DB CAGCTGCAGTGCAGAACAGAGAGTTTGCATCAGCTCCTCGAAGATGAGAAAGATGAGTGA 787  
 QY 181 GlnLysLeuGlnAsnIlelelaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
 DB CAAAATTTACAAATATCATAGCAGCGGGCTACTCAGTATAATCATGATGTAAGAGG 847  
 QY 201 LysGluArgGluTyrAsnLysLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
 DB AAGGAGCGGAATATAATGAAGCTAAGAGGAGCGCTGCATCAGCTCGTTATGAACAAGAG 907  
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
 DB GATAAAACATAGCCATGATGTTTAAATTTATGTGGTCGAGCTGATGGCAACAGAGC 967  
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
 DB TCATGGAGGACTGACAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATTTCTGTG 1027  
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
 DB AATGATTTATGATGACCGCAGAGAGAGATCCCTGATGTGAAGCCGCGAGCTGAAGAGGTC 1087  
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
 DB CTCACAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1147  
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
 DB GAAAGAGCAGAGAGACCGCACAGGCACTGTTGCTATCTCGATATAGAAGATGACTCTGG 1207  
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluLeuThr 340  
 DB GAACGTGAGCAGAGACAGCGGTGGGGCCCTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1267  
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
 DB AACACATCAGGAACACAGTGGAGATTTTGAAGTCAATGAGAAAACATCGATNAACCA 1327  
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380  
 DB GCTTCGAAGGTGCACTCAGAGGGCCCTTAATGAGGAGGAGCGTCACTCTCAGCAACAGACCAT 1387  
 QY 381 GluGlnGlnThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
 DB GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAAGAAATGATCAAGGCT 1447  
 QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
 DB CAGCAGCAGCTCTTACACAGCAGCTGCCACCGCTGATGATGATGACACCACTCTACTG 1507  
 QY 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLysGluArgLysGluGluTrpThrLeuPhe 440  
 DB TTGCAGACTGTTACTTACTTGAAGAAAGAGAACGCCCTTAAAGAGAGGTGGACCCCTTTT 1567  
 QY 441 LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaIleArg 460  
 DB AAAGAGCAAAAGAAATTTTGAAGAGAAAGCGGAGCTTTTACAGAAAGCTGCCATTCGA 1627  
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480  
 DB TTGGGGTTGGAGAGAAAGCGGTTTGAAGAGAGCGGAGCGCTGGTGAAGCAGCAGTTT 1687  
 QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500  
 DB TTAACATGACGAACCTTTGACCACCAGAACTCAGAAATGTGAAACTTTTTCAGTGCCTTC 1747

QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
 DB TCAGGAAGTCTTGATCCAGACAATCTTATAGTCCACTCACGCCACGGCAAGAGCTA 1807  
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540  
 DB CACAGTGTGGCTAATGGGGTGCCAGCTTGACATCAAACTGCTAAATCTCTTCTCTGCC 1867  
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560  
 DB TCACCTTCTACTTACAGCTTTTCGCCAGACACATTCATGTGTCTGTGACACAGTTCATC 1927  
 QY 561 SerValLeuAsnIleThrProGluLysSerLysProSerGluValAlaArgGluSerThr 580  
 DB AGTGTGCTGAATATTAACCTCTCAAGAAAGTAAACCAAGTGAAGTTGCAAGAGAAAGCAG 1987  
 QY 581 AspGlnLysTyrSerValGlnSerArgProSerArgGluGlyCysTyrSerGlyCys 600  
 DB GATCAGAAGTGGAGGCTGAGTCCGAGGCCAGCTCCGGAGGGGTGCTTACAGTGGATGT 2047  
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615  
 DB TCCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 2092  
 RESULT 5  
 BC078687  
 LOCUS  
 DEFINITION  
 Rattus norvegicus synovial sarcoma, X breakpoint 2 interacting protein, mRNA (cdna clone MGC:93056 IMAGE:7115887), complete cds.  
 ACCESSION  
 BC078687  
 VERSION  
 BC078687.1 GI:50926757  
 KEYWORDS  
 MGC.  
 SOURCE  
 Rattus norvegicus (Norway rat)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Rattus.  
 REFERENCE  
 1 (bases 1 to 2547)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
 Mammalian Gene Collection Program Team  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 1477932  
 2 (bases 1 to 2547)  
 NIH MGC Project  
 Direct Submission  
 Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Howard Jacobs  
 cDNA Library Preparation: Express Genomics  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Alignment Scores:		
Pred. No.:	9.62e-301	2547
Score:	2937.00	567
Percent Similarity:	95.1%	Conservative: 18
Best Local Similarity:	92.2%	Mismatches: 28
Query Match:	92.8%	Indels: 2
DB:	6	Gaps: 1

US-10-644-084-2 (1-615) x BC078687 (1-2547)

Qy	1	MesG	Asp	Trp	Met	Thr	Val	Thr	Asp	Pro	Val	Leu	Cys	Thr	Glu	Asn	Leu	Ser	20			
Db	366	ATGGG	AAG	TGG	GAT	GACT	GT	TAC	AAG	TCC	AGT	TCT	TGT	GTA	CAG	AAA	CA	AAAA	TCTCTCT	425		
Qy	21	Gln	Tyr	Thr	Ser	Glu	Thr	Lys	Met	Ser	Pro	Ser	Ser	Leu	Tyr	Ser	Gln	Gln	Val	Leu	Cys	40
Db	426	CAATATAC	CTC	GAAA	CAAA	GAT	GTCTCC	CGT	CAAG	TTT	TAT	TACT	CGC	AGC	CAAG	TACT	GTGC					485
Qy	41	Ser	Ser	Val	Pro	Leu	Ser	Lys	Asn	Val	His	Gly	Val	Phe	Cys	Thr	Gly					60
Db	486	CTGTGCA	ACAC	CTTTAT	TAT	CCA	AAG	ATG	CGAT	TGG	TTTTCT	CAG	TGC	CTTTCT	GC	CAGG	AGAG					545
Qy	61	Asn	Ile	Gln	Ser	Ile	Ser	Tyr	Leu	Asp	Gln	Glu	Leu	Thr	Thr	Phe	Gly	Phe	Pro	Ser		80

```

QY      441  LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db      1686  AAGAGCAAAAGAAATTTTGGAGAGAGAAAGACGAAGCTTTACAGAAGCTCTATTCCG 1745

QY      461  LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db      1746  TTGGGGTTGGAGAGAAAGCGTTTGGAGNAGAGCAGCAGCTGGTGAAGCAGCAGTTT 1805

QY      481  LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db      1806  TTAACATGACGACGACCTTTGATCACCAAGAACTCAGAAAATGTGAAACTTTTCAGTCCCTT 1865

QY      501  SerGlySerSerAspProAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db      1866  TCAGGAAGTTCTGATCCAGACAACTTATAGTCCACCCACGCGCCACGCAAAAGAAGCCA 1925

QY      521  HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db      1926  CACAGTGTCTGATNAGGGGTACACGCTTGACATCAAACTGGCTAAGTCTCTCCGACC 1985

QY      541  SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db      1986  TCACCT-----TCAGACTTTCGCCGCTCTCGCTCATGTGTCTGTGAGCAGAGTCCCGTC 2039

QY      561  SerValLeuAsnIleThrProGluSerLysProSerGluValAlaArgGluSerThr 580
Db      2040  AGTGGCGTGACTGTGACTCTCTGAAGAAACCAACCGAAATGAGGTTTGGAGAGAAAGTAGC 2099

QY      581  AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600
Db      2100  GACCAAGAGTGAGCGTGTGTCCAGACCCAGCTCCCGGAGGGTGTCTACGGTGGATGC 2159

QY      601  SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db      2160  TCTCGGCTACACAAGTCCACGCTCGNACGAGATGACTTACCA 2204

RESULT 6
AF532970
LOCUS      Rattus norvegicus afadin- and alpha-actinin-binding protein ADIP
DEFINITION
ACCESSION  AF532970
VERSION     AF532970.1
KEYWORDS   AF532970.1  GI:27451840
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 3195)
Asada,M., Irie,K., Morimoto,K., Yamada,A., Ikeda,W., Takeuchi,M.
and Takai,Y.
ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized
at Cell-Cell Adherens Junctions
J. Biol. Chem. 278 (6), 4103-4111 (2003)
12446711
2 (bases 1 to 3195)
Asada,M., Irie,K. and Takai,Y.
Direct Submission
Submitted (26-JUL-2002) Department of Molecular Biology and
Biochemistry, Osaka University Graduate School of Medicine/Faculty
of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan
Location/Qualifiers
1. 3195
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
79..1920
/codon_start=1
/product="afadin- and alpha-actinin-binding protein ADIP"
/protein_id="AAO15016.1"

CDS

```

```

/db_xref="GI:27451841"
/translation="MGDMWTVDPLVTENKNLSQYTSGETKMSPSLSYQQVLCSATP
LSKNVGVSAFCTGENIEQISYLDQELTTFFPFSLYEESKSEAKRELGIALLNC
MNELLVLQKNLLAQESVETQNLKLGSDMDHLQCYAKLKEQLSEARREMSLQERDR
QLOCKNRNLHOLLKNEKEVOKLQNI TASRATQYNDVKREREYVKLKERLHOLVMN
KKDKIAMDVLNVYGRVDGKGSWRTOKTEARNEDEMYKILLNDYEVYKOILLNNAE
LKKVLQOMKEMISLLSPQKKPRERAEDSTGTVVISDVEDDAGELSDGSKVSLCSDT
VREGUTSIRKQWIRILSKSHVEKLDNQASKVHSEGFHEBDVISRQDHEQETKJLEUE
RKEMIKAQQLQQLATCDDDTTSLLRDCLYLLBEKERLKEBMSLFPKQKNPFERE
RRSTAAAIQGLLERKAFEBERASWVQQLFNTMTTFOHNSENVKLSFAFGSSDPDN
LIVHPRPROKKPHSVANGVPACTSKLAKSLPTSPDFCPSRSCVSEHSPVSALTVTPE
ETKNEVGRSTDDQKNSVVRPSPSSREGYCGCSSAYTSSHVERDDLP"

```

## ORIGIN

```

Alignment Scores:
Pred. No.:      1,32e-300      Length:      3195
Score:          2937.00        Matches:      567
Percent Similarity: 95.1%      Conservative: 18
Best Local Similarity: 92.2%    Mismatches:   28
Query Match:     92.8%        Indels:       2
DB:              6            Gaps:         1

US-10-644-084-2 (1-615) x AF532970 (1-3195)

QY      1  MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db      79  ATGGGAGATTGGATGACTGTTTACAGATCCAGTCTCTGTGTACAGAAAAAATAATCTCTCT 138

QY      21  GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db      139  CAATATACCTCAGAAAAAAGATGTCTCCGTCAAGTTTATATCTCGACAAGTACTGTGTC 198

QY      41  SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db      199  TCTGCAACACCTTTATCCAGAATGTGCATCGTGTTCAGTGCCTTCGACGAGAGAG 258

QY      61  AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrPheGlyPheProSer 80
Db      259  AACATCGAACAAGAGTATTTTCGTATCTTGATCAGGAACCTGACTACCTTCGGTTTCCTTCC 318

QY      81  LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db      319  TTGTATGAAGAATCCAAAAGTAAGAGCGAGAGGAGGAGTTAAGTATAGTTGCTCTCTG 378

QY      101  AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db      379  AACTGCATGATGAGCTGCTTGTCTTCAGCGGAAGAACCTTCCTGGCCACGGAAGCGTG 438

QY      121  GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db      439  GAGACACAGAATCTGAAGCTGGGCGAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA 498

QY      141  LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db      499  CTTAAGGAACAAGTTGGAGGCGCTCCAGCGAGAGATGATCAGCCTTCAGGAGAGACAGA 558

QY      161  GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
Db      559  CAGCTACAGTGCAAAAACAGGAATTTGCATCAGCTCTCTGAAAACAGGAAGAAGAGGTA 618

QY      181  GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db      619  CAAAATTTACAAAATATCATAGCCAGTCGGGCTACTCAGTATATATCATGATGTGAAGAGA 678

QY      201  LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db      679  AAGGAGCGGAGGTACAATAAATGAAGGAGCGCTCTGCATCAGCTTGTATTGAAACAAGAG 738

QY      221  AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db      739  GATAAAAATATAGCCATGGAGCTTTTAAATTTACGTGGGCGAGTGGATGGNAACGAGGC 798

QY      241  SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260

```



	D	b	TCTCGGAGCAGTGTAAAAACAGAAACCAGGAATCAAGATGAAATGGTAACAATAATTCTGCTG	858
	Qy	A	asnAspTyrgluTyArgGlnLysglNlleLeuMetGluAenAlaGluLeuLysLysVal	280
	D	b	AAATGATTATGATGACCCCGAGAACCATCTCTGCTGGAGAATCGGAGCTGTGAAGAAGTC	918
	Qy	L	euGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
	D	b	CTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCTCAAAAAGAAACCCAGA	978
	Qy	G	luArgAlaGluAspGlyThrGlyThrValalaIleSerAspIleGluAspAspSergly	320
	D	b	GAAGAAGCAGAGGACAGCACAGGCACCTGTTGTCTACTCCGATGTAGAAGACGAGCTGGG	1038
	Qy	G	luLeuSerArgaspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
	D	b	GAGCTGAGCAGATGGTGTTGTGAGGCTTTCCTGTGCACCTGTCAAGGAGAGCTTTACA	1098
	Qy	A	snSerIleArGLysGlnTrpArgileLeuleuLysSerHisValGluLysLeuAspAsnGln	360
	D	b	AACAGCATCAGGAAGCAGTGGAGAAATCTGPAARAGCCATGTGGAANAAC TTGATAACCAA	1158
	Qy	A	laSerLysValHieSerGluGlyLeuAenGluAspValIleSerArgGlnAspHis	380
	D	b	GC TTTCAAAGGTACACTCAGAGGGCTTTCATGAAGAGGACGTCATCTCACGACAAAGACCAT	1218
	Qy	G	luGlnGluThrGluLysLeuGluLyuLueGluileGluArgCysLysGluMetIleLysAla	400
	D	b	GAGCAAGAGACTGGAACACTGGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT	1278
	Qy	G	lnGlnGlnLeuLeuGlnGlnLeuAlaThrCysAspAspThrThrSerLeu	420
	D	b	CAGCAGCAGCTCTTACAGCAACAGCTGGCCA CTGCGTGTGATGACGACACCACCTCACTG	1338
	Qy	L	euArgAspCysTyrl euLeuGluGluLysGluArgLeuLysGluLuTrpThrLeuPhe	440
	D	b	TTGCGAGACTGTACTTCTGTTGAAGAAAAGGACGCCCTTAAAGAAGAGTAGTGCTCCCTTTTT	1398
	Qy	L	ysGluGlnLysLysAenPheGluArgGluArgSerPheThrGluAlaAlaIleArg	460
	D	b	AAAGAGCAAAAAAAAAAGATTTTGGAGAGAGAAAGACGACCTTACAGAAAGCTGTATTCGC	1458
	Qy	L	euGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnInPhe	480
	D	b	TTGGGTTGGAGAGANAGGCGTTTGAGGAAGAGCGAGCCAGCTGGGTGAAGCAGCAGTTT	1518
	Qy	L	euAenMetThrAenPheAsphiGlnAenSerGluAenVallysLeuPheSerAlaPhe	500
	D	b	TTAAACATGACGACCTTTGATCACAGAACTCAGAAAATGTGAAACTTTTTCAGTGCGCTTT	1578
	Qy	S	erGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnLysLeuLeu	520
	D	b	TCAGGAAGTTCTGATTCAGACAATCTTATAGTCCACCCCAGGCCACGGCAAAAGAAGCCA	1638
	Qy	H	isSerValAlaAenGlyValproAlaCysThrSerLysLeuThrLysSerLeuproAla	540
	D	b	CACAGTGTCGTANTGGGTGCCAGCTTGCACTCCAACTGGCTAAGTCTCTCCGACC	1698
	Qy	S	erProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHuIsSerSerIle	560
	D	b	TCACCT-----TCAGACTTCTGCCCGTCTCGCTCATGTGTGTCTGAGCACAGTCCCGCTC	1752
	Qy	S	erValLeuAenIleThrProGluGluSerLysProSerGluValAlaAargGluSerThr	580
	D	b	AGTCGCTGTACTGTGACTCTCTGAAGAACCAGAAATGAGGTGGAAGAGAAAGTACG	1812
	Qy	A	splnLysTyrsTpSerValGlnSerArgProSerSerArgGluGlyCysTyrsSerGlyCys	600
	D	b	GACCAGAACTGACGCTGGTGTCCAGACCACAGCTCCCCGGAGGGTGTCTACGGTGGATGC	1872
	Qy	S	erSerAlaPheA-rgSerAlaHiGlyAspArgAspLeuPro	615

Db 1873 TCCTCGGCTACACAAAGCTCCCAAGTGGGACGAGATGACTTACCA 1917

RESULT 7  
BC033637

LOCUS BC033637 3220 bp mRNA linear PRI 27-JAN-2004

DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein, mRNA (CDNA clone MGC:45118 IMAGE:5578582), complete cds.

ACCESSION BC033637

VERSION BC033637.1 GI:21707136

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3220)

AUTHORS Strausberg R.L., Feingold, E.A., Grouse, L.H., Dejeu, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altshuler, S.F., Jordan, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Zeng, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ussdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McSwain, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

PUBMED 12477932

REFERENCE 2 (bases 1 to 3220)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NIHC)  
Gaithersburg, Maryland;  
Web site: <http://www.nisec.nih.gov/>  
Contact: [nisc\\_mgc@hgrc.nih.gov](mailto:nisc_mgc@hgrc.nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Strantrop, S., Thomas, P.J., Touchman, J.W., Teague, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 69 Row: 0 Column: 9  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7662381.  
Location/Qualifiers  
1. .3220  
/organism="Homo sapiens"



```

/dbl_type="mRNA"
/dbl_xref="taxon:9606"
/clone="MGC:45118 IMAGE:5578582"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
1..3220
/gene="SSX2IP"
/note="SSX2IP"
/note="Synonyms: KIAA0923, ADIP"
/db_xref="GeneID:117178"
152..1996
/gene="SSX2IP"
/codon_start=1
/product="SSX2IP protein"
/protein_id="AAH3637.1"
/db_xref="GI:21707137"
/db_xref="GeneID:117178"
/translaton="MGDMVTDPGLSSSEKTIISOYTSKTPSSLSYQVQLCSSIP
LSKNVHSPFSAPCTEDNIEQISLYLDQBLTTFGPSPLYEESKGRKRELNIIVAVLNC
MNELLYLORKNLQAENVETQNLKGLSDMHQSCYSKLEQLETSRREMIGLOQRDR
QLQCKRNHLQKNEDEVQKLNIIASRATQYNHDMKRKEREYNKLEKRLHQLVNN
KDKKIAMDIINLGRADGRGSRWRTGTEARNEDEMYKILINDYVRKQILMENA
LKKVLQMKXEMISLLSPKKPRRVDSTGTVISDVEEDAGELSRSEMDLSCETV
REOLTNSIRKOWRIKLSHVEKLDNOVSKVHLEGFNDEDIISRODHEORTEKLELQ
CKEMITQOQLLOQLATAYDDDTLSLRDCYLLBEKRLKEWSLFRKQKNFERER
RSFTAAIRLGLERKAFBEERASWLKQQLNMTTFDQNSENVKLFSPAGSSDMDNL
IVHSRQPKPHSVNSGSPVCMKSLTKSLPASPTSDFCQTRSCISEHSSINVLNITA
BEIKPNQVGGECTNQKWSVASRPGSQEGCYSGCSLSYTNHVEKDLIP"

```

gene

CDS

ORIGIN

Alignment Scores:

```

Pred. No.: 1 25e-284 Length: 3220
Score: 2786.50 Matches: 538
Percent Similarity: 93.2% Conservative: 35
Best Local Similarity: 87.5% Mismatches: 41
Query Match: 88.0% Indels: 1
DB: 5 Gaps: 1

```

US-10-644-084-2 (1-615) x BC033637 (1-3220)

```

Qy 1 MetGlyAspTyrMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 152 ATGGAGATTGGATGACTTTACAGATCCAGGTCTGTCTTCAGAAAGCAAACTATCTCT 211
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 212 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATATCTACACCAAGTGTATGT 271
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 272 TCTTCAATACCTTATATCGAAAATGTGCACAGTTTTCAGTGCCTTCCTGCACAGAAGAT 331
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80
Db 332 AATATTGACAGAGTATCTATATCTTGATCAGGAATGACTACTTTTGGTTTTTCCTTCA 391
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 392 TTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGAGATTTAAATATAGTAGTGTACTA 451
Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db 452 AATTGTATGAATAGCTGCTTGTGCTTCAGGGGAGAACCTTCTAGCTCAGGAAAATGTG 511
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db 512 GAGACACAGAAATTTGAAGCTGGGAAGTATGATGACCACTACAGAGCTGCTACTCAAA 571
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 572 CTTAAGGAACAACCTGGAAACCTCCAGGAGGGAAATGATTGGGCTTCAGGAAGAGACAGA 631

```

```

Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluVal 180
Db 632 CAGTTTACAATGTGAAGAACAGGAATTTGCATCAGCTACTAAAGAAATGAGAAGATGAGGTG 691
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 692 CAAAAAATTTACAAAATATATCATTTGCAAGTGCAGCTACTCAGTATATCATGATATGAAGAGA 751
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 752 AAAGAGCGTGAATATAATAAACTGAAGGAACGCTCTACATCAACTTGTATTGAACAGAAA 811
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 812 GATAAGAAAAATAGCTATGACATTTTGAATATATGTCCGGGAGAGCTGATGGAAAAAGAGC 871
Qy 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db 872 TCCTGGAGGACTGGTAAAAACTGAAAGCCAGGAATGAAGATGAATGTATATAAATTTCTCTG 931
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db 932 AATGATTATGAATATCGTCAGAAAAACAATCCTAATGGAAAAATGCAGAACTTTAAGAAGTT 991
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db 992 CTTCAACAAATGAAAAAGGAATGATTTCTCTCTCTCTCCCAAAAGAAAGAACTAGATA 1051
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
Db 1052 GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCCGATGTTGAAGAAGATCCCGGG 1108
Qy 321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1109 GAACCTAAGCAGAGAGATGATGTGGAGCCTTCTCTCTGTGAACCTGTGAGAGAGCAGCTTACA 1168
Qy 341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db 1169 AACAGCATCAGAAACACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAGCTTGATAACCAA 1228
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db 1229 GTTTCAAAGGTACACCTGGAAAGGTTTTAATGAAGATGTAATCTCACGACAGACCAT 1288
Qy 381 GluGlnGluThrGluLysLysLeuGluGluIleGluArgCysLysLysGluMetIleLysAla 400
Db 1289 GAACAAGAACTGAAAAAATCGAGTTAGAAATTCAGCAGCTGTAAAGAAATGATTAATAACT 1348
Qy 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1349 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATACCATTCACATA 1408
Qy 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluLysGluLysGluThrLeuPhe 440
Db 1409 TTACGAGACTGTTATTTGTTGGAAAGAAAGAGAGAGCTCTCAAAGAAAGATGTCCTTTT 1469
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
Db 1469 AAAGAGCAGAAAAAGAAATTTGAGGGGAGAGAGCAAGCTTTACAGAGCCGCTATTTCGC 1528
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1529 CTGGGATGGAGAGAAAGGCAATTTGAAGAGAAAGAGAGCCAGTTGCTGTTAAAGCAGCAGTT 1588
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1589 CTAATATGACTTACCTTTGACCACCAAGCTCAGAAATGTGAAACCTTTTCAGTGCCTTC 1648
Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1649 TCAGGAAGTCTGATGGGACAATCTTATAGTCAGCTCAGGAGCCGCCCAAGAGAGCCT 1708
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540

```



```
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnPhe 480
Db 1642 CTGGGATGGAGAGAAAGCAATTTGAAGAAGAGCCAGTTGGTTAAAGCAGCAGTTT 1701
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1702 CTAATAATGACTACCTTTTGACCACACAGAACTCAGAAATGTGAACATTTTCAGTCCCTTC 1761
Qy 501 SerGlySerSerAspProAspAsnLeuLeuValHisSerArgProArgGlnLysLysLeu 520
Db 1762 TCAGGAAGTCTTGATTTGGGACAACTCTTATAGTCACCTCAGGACGCCGCAAAAGAGCCT 1821
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1822 CACAGTGTCTAATGGGTCTCCAGTTTGCATGCTTAACTTACCTTAAATCTCTTCCCTGCT 1881
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1882 TCACCTTCCACTTCAGACTTTTTCAGACACAGCTTCTCGCATATCTGAACATAGTTCAATC 1941
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
Db 1942 AATGTAAGTGAATATATACCTGCTGAAGAAATTTAAACCAATCAGTTGGAGGAGAACGTACA 2001
Qy 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrrSerGlyCys 600
Db 2002 AATCAAAATGGAGTGGCGTCAAGACCTTGGATCACAGGAAGGTTGCTAGTAGTGATGC 2061
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2062 TCCTTGACGTACACAAATCTCATGTAGAGAAAGATGACTTACCT 2106
```

```
RESULT 9
AB023140
LOCUS Homo sapiens mRNA for KIAA0923 protein, partial cds. PRI 10-JAN-2004
DEFINITION Homo sapiens mRNA for KIAA0923 protein, partial cds.
ACCESSION AB023140
VERSION AB023140.1 GI:4589477
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1
Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirose, M.,
Miya, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
Prediction of the coding sequences of unidentified human genes.
XIII. The complete sequences of 100 new cDNA clones from brain
which code for large proteins in vitro
DNA Res. 6 (1), 63-70 (1999)
2 (bases 1 to 5835)
Ohara, O., Nagase, T. and Kikuno, R.
Direct Submission
Submitted (04-FEB-1999) Osamu Ohara, Kazuo DNA Research Institute,
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail: cdnainfo@kazuo.or.jp, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)
Location/Qualifiers
1..5835
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hh02882"
/sex="male"
/tissue_type="brain"
/clone_lib="pbluescriptII SK plus"
/dev_stage="adult"
1..5835
/gene="KIAA0923"
<229..2109
```

```
/gene="KIAA0923"
/note="Start codon is not identified"
/codon_start=1
/product="KIAA0923 protein"
/protein_id="BAA76767.2"
/db_xref="GI:40789987"
/translation="TSSSGILALEIAEAMGDWMTVTPDGLSSSEKTIISQVTSYSEKMPSS
LYSOVLCSIPLSKNVHSFFSAFCTEDNISQSIYSYLOELTTFGPFSLYSEKSGKET
KRENIIVAVLNCMNEILLVQRNLLAQENVTOKLQSGDMHDLQSCYSKLEQLETS
RREMIQJQRDRQLOCKNRNIIHQLEKEDVQKQNI IASRATQYNDHMKRKBREYN
KLKRLHQLVNNKKDKITAMDLNVVGRADGKRGSRGTKEARNEDEMYKILLNDYE
YRQKILMENAELKKVLQOMKEMISLSPKPKKPRVRVDDSTGTVI SDVEEDAGELS
RESMDLSCETVRSQLTNSIRKQRIILKSHVKEKLDNQVSKVHLEGFNEDVISRODHE
QETSEKLEIOOCCKEMIKTQQLQATAYDDDTTSLRDCYLLSEKRLKSEWSL
PKEQKQNFERRRSTAAIRGLERKAFEEERASWLKQQLFNMTTFPHQNSKNSKLF
SAFGSDMDNLIIVHSRQPKPHSVNSGSPVCMKSLTKSLPASPSDFQOTRSCIS
EHSINVLNITABEIKPNQVGECTNQKWSVASRPGSQGCVSGCSLSTYTNHVEKDD
Lp"
ORIGIN
Alignment Scores:
Pred. No.: 2,88e-284 Length: 5835
Score: 2786.50 Matches: 538
Percent Similarity: 93.2% Conservative: 35
Best Local Similarity: 87.5% Mismatches: 41
Query Match: 88.0% Indels: 1
DB: 5 Gaps: 1
US-10-644-084-2 (1-615) x AB023140 (1-5835)
Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 265 ATGGGAGATTGGATGACTGTTTACAGATCCAGGTCTGTCTTCAGAAAGCAAACTATCTCT 324
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLysGlnGlnValLeuCys 40
Db 325 CAATATACCTCAGAAACAAAGATGCTCCATCAGTTTATATCTCAGCAAGTGTCTATGT 384
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 385 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTCGCTTCGACAGAAAGAT 444
Qy 61 AsnIleGluGlnSerIleSerTyrTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 445 AATATTGAACAGAGTATCTCATATCTTGTATCAGAAATTGACTACTTTTGGTTTCTCTTCA 504
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 505 TTATATGAGNATCCAAAGGTAAAGAGACAAAGAGAGAGTTAAATATATAGTAGTGTACTA 564
Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db 565 AATTGTATGAATGAGTGTCTTGTCTTCAGCGGAAGAACCTTCTAGCTCAGGAAAAATGTG 624
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
Db 625 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGACCATCTTACAGAGCTGTCTACTCAA 684
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 685 CTNAGGAACACTGGAAACCTCCAGGAGGAAATGATTGGGCTTCAGAAAGAGACAGA 744
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
Db 745 CAGTTACAATGTAAGACACAGAAATTTTCATCAGCTACCTACTAAAGAAATGAGAAAGATGAG 804
Qy 181 GlnLysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 805 CAAAAAATACAAAATATCATTTGCAAGTCGAGCTACTCAGTATATATCATATGATGAAGA 864
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 865 AAGAGAGCGTGAATATATAATAAATGAGAAACGCTTACTCACTTGTATGTAACAGAA 924
```

```
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db |||||
QY 925 GATTAAGAAAATAGCTATGTCATTTTGAATATATGTCGGAGAGCTGATGGAAGAGGC 984
Db |||||
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db |||||
QY 985 TCCTGGAGGACTGGTAAACCTGAAGCCAGGAAATGAAGATGAATGATATAAATTTCTCTTG 1044
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db |||||
QY 1045 AATGATTATGAATATCGTCAGAAAACAAATCCTTAATGGAATAATGCAGAACTTAAAGAGTT 1104
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db |||||
QY 1105 CTTCAACAAATGAAGAAAGAAATGATTTCTCTTCTCCCAAGAAAGAAACCTAGA 1164
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db |||||
QY 1165 GAAAGAGTAGATGATAGTACAGGAACGTGTT--ATTTCGATGTTGAAGAAGATGCCGGG 1221
QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db |||||
QY 1222 GAACCTAAGCAGAGAGATGATGGGACCTTTCTCTGTGAAACTGTGAGAGAGCAGCTTACA 1281
QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
Db |||||
QY 1282 AACAGCATCAAGAAACAGTGGAGAAATTTGAAAGTCAATGTAGAAAAGCTTGATAACCAA 1341
QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380
Db |||||
QY 1342 GTTTCNAAGGTACACCTGGAGGTTTAAATCATGATGAATGTAATCTCAGCACAGACCAT 1401
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysLysGluMetIleLysAla 400
Db |||||
QY 1402 GAACAAGAAACTGAAAAACTCGAGTTAGAAATTCAGCAGCTGTAAGAAAGATGATTAAGAACT 1461
QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db |||||
QY 1462 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCTATGATGATGATACCACTTCACTA 1521
QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluTyrThrLeuPhe 440
Db |||||
QY 1522 TTACGAGACTGTTATTTGTTGGAAGAAAGAACGCTCTCAAGAAAGATGGTCCCTTTT 1581
QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaIleArg 460
Db |||||
QY 1582 AAAGAGCAGAAAAAGAAATTTTGAGAGGAGAGAGCAAGCTTTACAGAAAGCCGCTATTTCG 1641
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaLaserTrpValLysGlnGlnPhe 480
Db |||||
QY 1642 CTGGATTGGAGAGAAAGGCATTTGAGAGAGAAAGAGCCAGTGGTAAAGCAGCAGTTT 1701
QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db |||||
QY 1702 CTAAATATGACTACTTTTGACCACCAAGACTCAGAAAATGTGAAACTTTTCAGTGCCTTC 1761
QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db |||||
QY 1762 TCAGGAAGTCTGATTGGGACAATCTTATAGTGCACTCGAGGAGCCGCCAAAGAACCT 1821
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db |||||
QY 1822 CACAGTGTGCTAATAGGTGCTCCAGTTGCGATGCTAAACTTACTAAATCTCTTCTCTGCT 1881
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db |||||
QY 1882 TCACCTTCCACTTCAGACTTTTTCGACAGACAGCTTCTCGCATATCTGAACATAGTTCAATC 1941
QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
Db |||||
QY 1942 AATGTACTGAATATAACTGCTGAAGAAATTAACCAAAATCAGGTGGAGGAGAAATGACA 2001
```

```
QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
Db |||||
QY 2002 AATCAAAAATGGAGTGGCATCAAGACCTGGATCAGAGGAGTTCTCTAGTGGATGC 2061
QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db |||||
QY 2062 TCCTTGAGCTACAAAATCTCATGTAGAAAAAGATGACTTACCT 2106

RESULT 10
AY367055
LOCUS
DEFINITION
Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein
(SSX2IP) mRNA, complete cds.
ACCESSION
AY367055
VERSION
AY367055.1
KEYWORDS
GI:34451652
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 2379)
Lu, L., Huang, X.Y., Xu, M., Yin, L.L., Li, J.M., Zhou, Z.M. and Sha, J.H.
Cloning a new transcript of X breakpoint 2 interacting
protein (SSX2IP) in testis
Unpublished
2 (bases 1 to 2379)
Sha, J.H., Zhou, Z.M. and Li, J.M.
Direct Submission
JOURNAL
TITLES
Submitted (13-AUG-2003) Key Lab of Reproductive Medicine, Nanjing
Medical University, 140 Han Zhong Road, Nanjing, Jiangsu 210029,
China
FEATURES
Location/Qualifiers
1..2379
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/tissue_type="testis"
1..2379
/gene="SSX2IP"
/contig="190..2034"
/codon_start=1
/product="synovial sarcoma, X breakpoint 2 interacting
protein"
/protein_id="AA072373.1"
/db_xref="GI:34451653"
/translation="MGDMWTVTDPLSSESKTISQVTSKTPSSLYSQVLCSSIP
LSKNVHSFTSAFTEQNIQSISYLDQELTTFGPPSLYBSKGTREKLNIVAVLNC
MNELLVLRKNLLAQENVETONLKGSDMDHLQSCYKLEQLETSRREMIGLQERDR
QLQCKNRNLHQLLKNSEKDEQKLNLIASRATQYNHDMKREYNKLKERLHOLVNV
KDKK1AMDILNYAGRADRGSRWRTGKEARNEDEMYKILLNDYERQKILLMENA
LKVLQOQMKKEMIISLSPQKKPRPRVDDSTGTVISDVEDAGELSRSHWJLSCETV
REGLTISRIRKQWIRLKHVEKLDNQVSKVHLEGFNDEDVSRQDHEQETKLELEIQ
CKEMIKTQOOLQOQLATAYDDTSLLRDCYLLEKERLKEEMSLFKQKQKQFRER
RSFTEARILGLERKAPFEERASWLKQOFLNMTTFDHNSENKLFKFSGSSPDWNL
IVHSRQPKPKHVSNGSPVCMKSLTKSLPASPTSDFCQTRSCISPHSINVNLITA
EIKPNQGVGECTNQKRVASRFPQSGCGCYSGCSLSYTNHSHVKDDL P"

Alignment Scores:
Pred. No.: 2,76e-284 Length: 2379
Score: 2761.50 Matches: 537
Percent Similarity: 93.0% Conservative: 35
Best Local Similarity: 87.3% Mismatches: 42
Query Match: 87.9% Indels: 1
DB: 5 Gaps: 1

US-10-644-084-2 (1-615) x AY367055 (1-2379)

QY 1 MecGlyAspTrpMetThrValThrAppProValLeuCysThrGluAsnLysAsnLeuSer 20
|||||
```

```
Db 190 ATGGGAGATTGGATGACTGTTACAGATCCAGGTCTGTCTTCAGAAAGCAAACTATCTCT 249
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerLeuTyrSerGlnValLeuCys 40
Db 250 CAATATACCTCAGAAACAAGATGCTCCATCAAGTTTATATACACAGCAAGTGTATGT 309
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 310 TCTTCAATACCTTTATCGAAANAATGTGCACAGTCTTTTTCAGTGCCTTCTGCACAGAAGAT 369
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 370 AATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTTCCTTCA 429
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
Db 430 TTATATGAAGAAATCCAAAGGTAAGAGACAAGAGAGAGTTAAATATATAGTAGCTGTACTA 489
Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
Db 490 AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAANAATGTG 549
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnInsSerCysTyrAlaLys 140
Db 550 GAGACACAGAATTTGAAGCTCGGAAGTGAATGATGACCATCTACAGAGCTGTACTCAAAA 609
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
Db 610 CTTAAGGAACAACCTCGAAACCTCCAGGAGGGAATGATTGGCTTCAGGAAGAAGACAGA 669
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180
Db 670 CAGTTTACAAATGTAGAAACAGGAATTTGCTCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 729
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
Db 730 CAAAAATTACAAATATCATTCATTCGAACTCGAGCTACTCAGTATATATCATGATATGAAGAGA 789
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
Db 790 AAAGAGCGTGAAATATAATAAACTGAAGCAAGCGCTACATCAACTTGTATTGAACAAGAAA 849
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
Db 850 GATTAAGAAAATAGCTATGGAATTTTGAATTTATGCCGGGAGAGCTGTATGGAANAAGAGGC 909
Qy 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
Db 910 TCCTGGAGGACTGGTAAAACCTGAAGCCAGGAATGAAGATGAATGTATGAATTTCTCTTG 969
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
Db 970 AATGATTATGAATATCGTCAGAAAACAAATCCTAATGGAANAATGCAGAACTTTAAGAGGTT 1029
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
Db 1030 CTTCAACAATGAAAAAGGAATGATTCTCTTCTTCTCCCAAAAAGAAAAACCTTAGA 1089
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320
Db 1090 GAAAGAGTAGATAGTACAGGAACCTGTTT---ATTTCGATGTTGAAGAAGATGCCGGG 1146
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
Db 1147 GAACTAAGCAGAGAGTATGTGGACCTTTTCTCTGTGAAACTGTGAGAGAGCAGCTTACA 1206
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLysAspAsnGln 360
Db 1207 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAAGTCATGTAGAAAAAGCTTTGATAACCAA 1266
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGlnGluAspValIleSerArgGlnAspHis 380
Db 1267 GTTTCAAAGGTACACCTGGGAAGGTTTTTAATGATGAAGATGTAATCTCACGCGCAAGACCAT 1326
```

```
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla 400
Db 1327 GAAACAAGAAACTGAAAAAATCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAAAACT 1386
Qy 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1387 CAGCAACAGCTTTTACAGCAGCAGCTGCTACTGTCATATGATGATGATACCACTTCACATA 1446
Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
Db 1447 TTACGAGACTGTTATTTGTTGGAAGAAAGAAAGAAAGCTCTCAAGAAAGAAAGTGTCCCTTTT 1506
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg 460
Db 1507 AAAGAGCAGAAAAAGAAATTTTGAGAGGGAGAGAGACGAAGCTTTACAGAAGCCGCTATTGCG 1566
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1567 CTGGATTGGAGAGAAAGGCCATTTGAAGAAAGAAAGAGCCAGTTGGTTTAAAGCAGCAGTTT 1626
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1627 CTAAATATGACTACTCTTTGACCACAGAACTCAGAAAATGTGAAAATCTTTTCAGTCCCTTC 1686
Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1687 TCAGGAAGTTCTGATTGGGACAACTTTATATAGTCACCTCGAGCGCCGCAAAAAGAGCCT 1746
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1747 CACAGTGTGTAAATGGGCTCTCCAGTTTGCATGTCTAAACTTACTAAATCTCTTCTCTGCT 1806
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1807 TCACCTTCCACTTCAGACTTTTGCACAGACAGCTTCTGTCATATCTGAACATAGTTCAATC 1866
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaAArgGluSerThr 580
Db 1867 AATGTACTGAATATAACTGCTGAAGAAATTAACCAAAATCAGGTGGAGTAGAATGTACA 1926
Qy 581 AspGlnLysTrpSerValGlnInsSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
Db 1927 AATCAAAAATGGAGTGGGCATCAAGACCTTGATTCAGAGGAAGGTTGCTATAGTGGATGC 1986
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 1987 TCCTTGAGCTACAAATTTCTCATGTAGAAAAGATGACTTACCT 2031
```

## RESULT 11

```
AR718997
LOCUS AR718997 2272 bp DNA linear PAT 07-OCT-2005
DEFINITION Sequence 1497 from patent US 6943241.
ACCESSION AR718997
VERSION AR718997.1 GI:77367951
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2272)
AUTHORS Isegai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.-i., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masubo,Y.
TITLE Full-length cDNA
JOURNAL Patent: US 6943241-A 1497 13-SEP-2005;
Research Association for Biotechnology; Tokyo;
WOX;
FEATURES
source Location/Qualifiers
1..2272
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
```

Alignment Scores:

Pred. No.: 8,66e-278 Length: 2272  
Score: 2720.00 Matches: 536  
Percent Similarity: 86.8% Conservative: 34  
Best Local Similarity: 81.6% Mismatches: 40  
Query Match: 85.9% Indels: 47  
DB: 2 Gaps: 3

US-10-644-084-2 (1-615) x AR718997 (1-2272)

Qy 1 MetGlyAspTprMetThrValThrAspProValLeuCysThrGluAenLysAsnLeuSer 20  
Db 224 ATGGAGATTTGGATGACTGTGTACAGATCCA-----GAAACCAAACTATCTCT 271  
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Db 272 CAATATACCTCAGAAACAAAGATGCTCCATCAAGTTTATACCTCACAGCAAGTGTATGT 331  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 332 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTCAGTGCCTTCTCGCACAGAAGAT 391  
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
Db 392 AATATTGACACAGATATCTCATATCTTGTATCAGGAATTGACTACTTTTGGTTTCTCTTCA 451  
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAenIleValAlaValLeu 100  
Db 452 TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGTAAATATAGTAGCTGTACTA 511  
Qy 101 AsnCysMetAsnGluLeuValLeuValLeuGlnArgLysAsnLeuLeuAlaGlnLeuSerVal 120  
Db 512 AATTGTATGAATGAGCTGCTTGTCTTTCAGCGGAAGAACCTTCTAGCTCAGGAATGTG 571  
Qy 121 GluThrGlnAenLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db 572 GAGACACAGAATTTGAAGCGGGAAGTGATATGACCACCTACACAGAGCTGCTACTCAAAA 631  
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Db 632 CTTAAGGAACAACCTGGAACCTCCAGGAGGGAATGATTTGGCTTCAGGAAGAGACAGA 691  
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluVal 180  
Db 692 CAGTTACAATGTAAGACAGGAATTTGCTCAGCTACTAAGAAATGAGAAGATGAGGTTG 751  
Qy 181 GlnLysLeuGlnAenIleAlaSerArgAlaThrGlnTyrAenHisAspValLysArg 200  
Db 752 CAAAAATTACAAATATATCATTTGCAAGTCGACTACTCAGTATATCATGATATCAAGAGA 811  
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Db 812 AAAGAGCGTGAATATAATAACTGAAGAACCTCTACATCAACTTTGTTATGAACAAGAAA 871  
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db 872 GATAAGAAAAATAGCTATGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAGGC 931  
Qy 241 SerTprArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Db 932 TCCTGGAGGACTGGTAAACTGAAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 991  
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAenAlaGluLeuLysLysVal 280  
Db 992 AATGATTATGAATATCGTCAGAAAATAATCTTAATGGAATAATGCAAGAACTTAAGAGGTT 1051  
Qy 281 LeuGlnGlnMetLysLysGluMetLysSerLeuLeuSerProGlnLysLysLysProArg 300  
Db 1052 CTTCAACAAATGAAAAAGGAATGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1111  
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320

Db 1112 GAAAGAGTAGATGATAGTACAGGAACCTGTT--ATTTCGATGTTGAAGAAGATGCCGGG 1168  
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
Db 1169 GAACTAAGCAGACAGATAGTGTGGACCTTTCTGTGTGAACCTGTGAGAGAGACCTTACA 1228  
Qy 341 AsnSerIleArgLysGlnTprArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
Db 1229 AACAGCATCAGAAACAGAGTGGAGATTTTGAAGAAGTCATGTAGAAAAGCTTGTATACCAA 1288  
Qy 361 AlaSerLysValHisSerGluGlyLeuAenGluAspValIleSerArgGlnAspHis 380  
Db 1289 GTTTCAAAGGTACACCTGGAAGGTTTTAATGATGAAGATGAATCTCACGACAAGACCAT 1348  
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
Db 1349 GAACAAGAACTCGAAGTTAGAAATTCAGCAGCTGTAAAGAAAATGATTTAAAACT 1408  
Qy 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
Db 1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATATACCACTTCACTA 1468  
Qy 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluThrThrLeuPhe 440  
Db 1469 TTACGAGACTGTTATTTTGTGGAAAGAAAGAACGTCCTCAAGAAAGAAATGGTCCCTTTT 1528  
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaIleArg 460  
Db 1529 AAAGACGACAAAAAGAAATTTTGAGAGGAGAGACGAAGCTTTTACAGAAGCCGCTATTCCG 1588  
Qy 461 LeuGlyLeuGlu----- 464  
Db 1589 CTGGGATTTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACACAAGAAATCTCTTTGGGT 1648  
Qy 464 ----- 464  
Db 1649 ATCCATCATTCACCCAGGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1708  
Qy 465 -----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnG 479  
Db 1709 ATTTGCTGTTTAAAAAAGAAAGGCATTTTGAAGAAGAAAGAGCCAGTTGGTTAAAGCAGCA 1768  
Qy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499  
Db 1769 GTTCTTAATATGACTACCTTTGACCACCAAGAACTCAGAAAATGTGAAACTTTTCAGTGC 1828  
Qy 499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLy 519  
Db 1829 CTTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTGCACTCCGAGGCAGCCGCAAGAA 1888  
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539  
Db 1889 GCCTCAGAGTGTCTAATGGGTCTCCAGTTTCAGTGTCTAAACTTACTAAATCTCTCTCC 1948  
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559  
Db 1949 TGCTTACCTTCCACTTTCAGACTTTTCCAGACACACCTTCTGCATATCTCAACATAGTTC 2008  
Qy 559 rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579  
Db 2009 AATCAATGACTCAATATAAATCTGCTGAAGAAAATTAACCAAAATCAGGTTGGAGGAGAACG 2068  
Qy 579 rThrAspGlnLysTprSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGl 599  
Db 2069 TACAAATCAAAATGAGGTGTGGCTCAAGACCTGGATCATCAGGAAGGTTGCTATAGTGG 2128  
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615  
Db 2129 ATGCTCTCTGAGCTACACAAATCTCATGTAGAAAAAGATGACTTACCT 2177  
RESULT 12  
AX747972  
LOCUS

DEFINITION Sequence 1497 from Patent EPI308459.  
ACCESSION AX747972.1 GI:32132360  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1  
AUTHORS Isegaki, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R., Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahara, K. and Masuho, Y.  
TITLE Full-length cDNA sequences  
JOURNAL Patent: EP 1308459-A 1497 07-MAY-2003;  
Helix Research Institute (JP); Research Association for Biotechnology (JP)

FEATURES  
source  
1. .2272  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

ORIGIN  
Alignment Scores:  
Pred. No.: 8.66e-278 Length: 2272  
Score: 2720.00 Matches: 536  
Percent Similarity: 86.8% Conservative: 34  
Best Local Similarity: 81.6% Mismatches: 40  
Query Match: 85.9% Indels: 47  
DB: 2 Gaps: 3

US-10-644-084-2 (1-615) x AX747972 (1-2272)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
DB 224 ATGGAGATGGGATGCTGTACAGATCCA-----GAAAGCAGAAACTATCTCT 271

QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
DB 272 CAATATACCTCAGAACAAAGATGCTCCATCAAGTTTATATCTACAGCAAGTGTATGT 331

QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
DB 332 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAGAT 391

QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
DB 392 AATATTGAACAGAGATATCTCATATCTTGATCAGGAATTTGACTACTTTTGGTTTCTCTCA 451

QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
DB 452 TTATATGAGAGATCCAAAGGTAAAGAGACAAAGAGAGAGATTTAAATATATAGTGTACTA 511

QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
DB 512 AATTGTATGAATGAGCTGCTTGTGCTTTCAGCGGAGAGACCTTCTAGCTCAGGAAAATGTG 571

QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
DB 572 GAGACACAGAAATTTGAAGCGGAGAGTATGATGGACCATCTACAGAGCTGCTACTCAAAA 631

QY 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160  
DB 632 CTTAAGGAACAACTGGAACCTCCAGAGGAGAAATGATTTGGGCTTTCAGGAAGAGACAGA 691

QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180  
DB 692 CAGTTACATGTAAAGAACAGGAAATTTGCATCAGCTACTTAAGAAATGAGAAAGATGAGGTG 751

QY 181 GlnLysLeuGlnAsnIleLeuSerArgAlaThrGlnTyrAsnHisAspValLysArg 200



Qy 519 sLeuHisSerValalaenGlyValProAlaCysThrSerIysLeuThrIysSerLeuPr 539  
 Db 1889 GCCTCACAGTGTCTCAATGGGTCTCCAGTTTGCAATCTTAAACTTACTAAATCTCTTCC 1948  
 Qy 539 oAlaSerProSerThrSerAspPheArGlnThrHisSerCysValSerGluHisSerSe 559  
 Db 1949 TGCTTCCACTTCCACTTCCAGATTTTCCGACAGACAGTTCCTGCATATCTGACATATGTC 2008  
 Qy 559 rILeSerValLeuAenleThrProGluGluSerIysPheProSerGluValalaArgGluSe 579  
 Db 2009 AATCAATGTACTGAATATAACTGCTGAAGAAATTAACCAAAATCAGGTGCGAGAGAAGC 2068  
 Qy 579 rThrAspGlnIysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGl 599  
 Db 2069 TACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCAGAGGAAGTGTCTATATG 2128  
 Qy 599 YCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
 Db 2129 ATGCTCTTGAGCTACACAAATTCATGTAGAAAAGATGACTTACCT 2177  
 RESULT 13  
 AK093173  
 LOCUS Homo sapiens cDNA FLJ35854 fis, clone TESTI2007113. 2272 bp mRNA linear PRI 20-JAN-2006  
 DEFINITION  
 AK093173  
 ACCESSION  
 AK093173.1 GI:21751964  
 VERSION  
 oligo capping; fis (full insert sequence).  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1  
 Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R.,  
 Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H.,  
 Sekine, M., Oabayashi, M., Nishi, T., Shibahara, T., Tanaka, K.,  
 Iehi, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y.,  
 Nagahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M.,  
 Shiratori, A., Sudo, H., Hosiiri, T., Kaku, Y., Kodaira, H., Kondo, H.,  
 Sugawara, M., Takahashi, M., Kanda, K., Yokoi, T., Furuya, T.,  
 Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Katsuta, N., Sato, K.,  
 Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H.,  
 Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M.,  
 Hiraoaka, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S.,  
 Yoshida, M., Hoteuta, T., Kusano, J., Kanehori, K., Takahashi, F., Hara, R.,  
 Hara, H., Tanase, T. O., Nomura, Y., Togiya, S., Komai, F., Hara, R.,  
 Takeuchi, K., Arita, M., Imose, N., Musashino, K., Yuuki, H., Oshima, A.,  
 Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T.,  
 Shiohata, N., Sano, S., Mori, Y., Sano, S., Momiya, H., Satoh, N., Takami, S.,  
 Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H.,  
 Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T.,  
 Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K.,  
 Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M.,  
 Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y.,  
 Ozaki, K., Hirao, M., Omori, Y., Kawabata, A., Hikiji, T., Kobatake, N.,  
 Inagaki, H., Ikeda, Y., Okamoto, S., Okitani, R., Kawakami, T.,  
 Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K.,  
 Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togashi, T.,  
 Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J.,  
 Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K.,  
 Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R.,  
 Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Iso, G., Iso, G., Ison, S.  
 Complete sequencing and characterization of 21,243 full-length  
 human cDNAs  
 Nat. Genet. 36 (1), 40-45 (2004)  
 14702039  
 2  
 Iehibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,  
 Ono, Y., Hoteuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,  
 Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,  
 Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,  
 Wakamatsu, A., Iehi, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y.,

Saito, K., Nishikawa, T., Kimura, K., Matsuoka, K., Nakamura, Y.,  
 Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fujii, A.,  
 Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,  
 Nagahari, K., Masuho, Y., Nagai, K., and Ison, S.,  
 NEDO human cDNA sequencing project  
 Unpublished  
 3 (bases 1 to 2272)  
 Ison, S. and Yamamoto, J.  
 Direct Submission  
 Submitted (04-JUL-2002) Takao Ison, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: flj-cdna@nifty.com, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
 Research Association for Biotechnology (RAB); cDNA library  
 construction: Helix Research Institute (HRI) (supported by Japan  
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
 HRI, and Biotechnology Center, National Institute of Technology and  
 Evaluation; clone selection for full insert sequencing: HRI and  
 RAB; annotation: HRI and RAB.  
 FEATURES  
 Location/Qualifiers  
 1..2272  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="TESTI2007113"  
 /tissue\_type="testis"  
 /clone\_lib="TESTI2"  
 /note="cloning vector: pME18SFL3"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8 666-278 Length: 2272  
 Score: 2720.00 Matches: 536  
 Percent Similarity: 86.8% Conservative: 34  
 Best Local Similarity: 81.6% Mismatches: 40  
 Query Match: 85.9% Indels: 47  
 DB: 3 Gaps: 3  
 US-10-644-084-2 (1-615) x AK093173 (1-2272)  
 Qy 1 MetGlyAspTrpMetThrValThrAppProValLeuCysThrGluAenLysAenLeuSer 20  
 Db 224 ATGGGAGATTGGATGACTGTTCAGATCCA-----GAAAGCAAAACTCTCTCT 271  
 Qy 21 GlnTyrThrSerGluThrIysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
 Db 272 CAATATACCTCAGAAAACAAAGATGCTTCCATCAAGTTTATATACACGCAAGTCTATGT 331  
 Qy 41 SerSerValProLeuSerLysAenValHisGlyValPheGlyValPheCysThrGlyGlu 60  
 Db 332 TCTTCAATACCTTATCGAAAAATGTGCACAGTTTTTTCAGTGCCTTCTGCACAGAAGAT 391  
 Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
 Db 392 AATATTGAACACAGATGATCTCATATCTTGTATCAGAAATGACTACTTTTGGTTTTCTCTCA 451  
 Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAenLysValAlaValLeu 100  
 Db 452 TTATATGAAGAATTCGAAGTAAGAGACAAAGAGAGAGTTAATATAGTAGCTGTACTA 511  
 Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAenLeuLeuAlaGlnGluSerVal 120  
 Db 512 AATTGTATGAATGAGTGTCTTGTCTCAGCGGAAGAACCTTCTAGCTCAGGAAAATGTG 571  
 Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
 Db 572 GAGACACAGAAATTTGAAGCGGGAAGTATGATATGACCATCTACAGAGCTCTACTCAAAA 631  
 Qy 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 Db 632 CTTAAGGAACAACTGGAAAACCTCCAGAGGGGAATGATTGGCTTTCAGGAAGAGACAGA 691





Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: nisc.mc@nih.gov  
 Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
 Maduro, O.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 139 Row: k Column: 13  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 41281570.

#### FEATURES

source

1..2540  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:75026 IMAGE:5745142"  
 /tissue\_type="Brain, adult medulla"  
 /clone\_lib="NIH MGC\_119"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"

gene

1..2540  
 /gene="SSX2IP"  
 /note="synonyms: KIAA0923, ADIP"  
 /db\_xref="GeneID:117178"  
 /db\_xref="MIM:608690"

CDS

241..1920  
 /gene="SSX2IP"  
 /codon\_start=1  
 /product="SSX2IP protein"  
 /protein\_id="AAH64389.1"  
 /db\_xref="GI:39963670"  
 /db\_xref="GeneID:117178"  
 /db\_xref="MIM:608690"  
 /translation="MGDMVTVDPLSSSEKTSIQYSETMSPSSLYSQVLCSSIP  
 LSKNVHFFAFCFEDNIQSISYLDQELTFPFSLYESKGETKRELNIVAVLNC  
 MNELVLRKNLLAQENVETQNLKGLMDHLQSCYKLEOLETSRREMIGLQERDR  
 QLOCKRNHLQNKEDVQKLNIIASRATQYNHDKERKEREYNKLERLHQLVNN  
 KDKKIAMDILNYYGRADGKGSWRTGKTARNEDVMYKILLNDYEVROKQILWENAE  
 LKKVLOMKKEMISLLSPQKKPRVDDSTGTVISDVEADGELSRESMDLSCETV  
 REQUTNIRKQWRILKSHVEKLDNQSVKHLEGNEDVDVSRQDHEQETKLEIQO  
 CKEMIKTQQLQOLATAYDDDTSLLRDCYLLEERKLEKESLFPKEQKNFERRR  
 SFTBAALRLGLRKAFAFEERASWLKQQLFMNTTTHQNSKNVSLFSAFSGSSDNDLI  
 VHSRQPKKHSHVNSGSPVCMKSLTKSLPASPSSTDFCQTRSCISEHAKN"

#### ORIGIN

Alignment Scores:

Pred. No.:	1..21e-274	Length:	2540
Score:	2691.00	Matches:	536
Percent Similarity:	80.6%	Conservative:	35
Best Local Similarity:	75.7%	Mismatches:	42
Query Match:	85.0%	Indels:	96
DB:	5	Gaps:	3

US-10-644-084-2 (1-615) x BC064389 (1-2540)

Qy	1	MetGlyAspTTPMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	241	ATGGAGATTGGATGACTGTTACAGATCCAGGTCTGTCTTCAGAAAGCAAACTATCTCT	300
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys	40

Db	301	CAATATACCTCAGAAAAAAGAGATGCTCCATCAAGTTTATATACACAGCAAGTCTATGT	360
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	361	TCCTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTTCAGTGCCTTCTGCACAGAAGAT	420
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	421	AATATTGACAGAGATATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTCTCTTCA	480
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu	100
Db	481	TTATATGAGAATCCCAAGGTAAGAGACAAAGAGAGAGTTAAATATATAGTACTGTTACTA	540
Qy	101	AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal	120
Db	541	AATTTGATGATGAGTGTCTTGCTTCAGCGGAAGAACCTCTAGCTCAGGAATAATGTG	600
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140
Db	601	GAGACACAGATTTGAAGCTGGGAAGTGATATCGACCATCTACAGAGCTCTACTCAAAA	660
Qy	141	LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg	160
Db	661	CTTAAGGAACAACCTCGGAACCTCCAGGAGGAAATGATTGGCTTCAGGAAGAGACAGA	720
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal	180
Db	721	CAGTTACAATGTAAGAACAGGAATTTGTCATCAGCTACTAAGAATGAGAAGATGAGGTG	780
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg	200
Db	781	CAAAAATTACAAAATATCATTTGCAAGTCGAGCTACTCAGTATATATATAAATCTCTTG	840
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys	220
Db	841	AAAGAGCGTGAATATATAAATGAACTGAAGGAACGCTACATCAACTTGTATGAAACAAGAA	900
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240
Db	901	GATAAGAAAATAGCTATGGACATTTTGAATATATGTCGGGAGAGCTGATGGAAGAGGCG	960
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
Db	961	TCCTGGAGGACTGCTGTAAGCCAGGAATGAAAGTGAATGATAAATAATCTCTTG	1020
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280
Db	1021	AATGATTTAGAAATATCGTCAGAAACAATCTTAATGAAATGCAAGAACTTAAGAAGGTT	1080
Qy	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
Db	1081	CTTCAACAATGAAAAGGAAATGATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1140
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320
Db	1141	GAAAGAGTAGATGATAGTACAGGAACCTGTT--ATTTCGATGCTGAAGAAGATGCCGGG	1197
Qy	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
Db	1198	GAACTAAGCAGAGAGATGATGTCGGGACCTTTCTCTGTAAGTGTGAGAGAGAGCTTACA	1257
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln	360
Db	1258	AACAGCATCAGAAACAGTGGAGAAATTTTGAAGATCATGTAGAAAAGCTTGTATACCAA	1317
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis	380
Db	1318	GTTCACAAAGGTACACCTGGGAAGGTTTTTAATGATGAAGATGTAATCTCACGACAGACAT	1377
Qy	381	GluGlnGluThrGluLysLeuGluLeuGluLeuGluArgCysLysGluMetIleLysAla	400
Db	1378	GAAACAAGAAACTGAAAACTCGAGTTAGAAATTCAGCAGGTGTAAAGAAATGATTAAACT	1437



```
Db 426 CAGTATAATCATGATATGAAGAGAAAAGAGCGTGAATATAATRAACTGAAGAACGCTTA 485
Qy 213 HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
|||
Db 486 CATCAACTGTTTATGAACAAGAAAGATGAAGAAATAGCTATGGACATTTTGAATATATGTC 545
|||
Qy 233 GIVArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaAArgAsnGlu 252
|||
Db 546 GGGAGAGCTGATGAAAAGAGGCTCTCTGGAGGACTCGTAAACTGAAGCCAGGAATGAA 605
|||
Qy 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
|||
Db 606 GATCAATGTATATAAATTCCTTGAATGATTATGAATATCGTCAGAAAACAAATCCTAATG 665
|||
Qy 273 GluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeu 292
|||
Db 666 GAAAATGCAGAACTTAAGAGAGTTCTTCAACAAATGAAAAGGAATGATTTCTCTCTCTT 725
|||
Qy 293 SerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312
|||
Db 726 TCTCCCAAAAAGAAAGAACCTTAGAGAAAGAGTAGATGATAGTACAGGAACGTGT---ATT 782
|||
Qy 313 SerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332
|||
Db 783 TCCGATGTTGAAGAAGATGCCGGGGAACCTAAGCAGAGAGAGTAGTATGGGACCTTTTCCTGT 842
|||
Qy 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
|||
Db 843 GAAACTGTGAGAGAGCAGCTTTACAACAGCATCAGAAAACAGATGGAGAAATTTTGAAAAGT 902
|||
Qy 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372
|||
Db 903 CATGTAGAAAAGCTTGATACCAAGTTTCAAGGTACACCTGGGAAGGTTTTTAATGATGAA 962
|||
Qy 373 AspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLeuGluIleGlu 392
|||
Db 963 GATGTAATCTCACGACAGACCATGNAACAAGAAACTGAAAACTCGAGTTAGAAATTCAG 1022
|||
Qy 393 ArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThr 412
|||
Db 1023 CAGTGTAAAGAAATGATTAATAACTCAGCAACAGCTTTTAA---CAGCAGCTCGCTACTGCA 1079
|||
Qy 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArg 432
|||
Db 1080 TATGATGATGATACACCTTCACTATTACGACCTGTTATTGTTGGAGAGAAAAGAACGT 1139
|||
Qy 433 LeuLysGluGluTyrThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArg 452
|||
Db 1140 CTCAAAGAGAAGATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGGGGAGAGACGA 1199
|||
Qy 453 SerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg 472
|||
Db 1200 AGCTTTTACAGAAAGCCGCTATTTCGCTGGGATTGGAGAGAAAGGCATTTGAAGAAGAAGA 1259
|||
Qy 473 AlaSerTrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
|||
Db 1260 GCCAGTTGGTTAAGCAGCAGCTTTCTAATATAGTACTACTTTTGACCACCAGAACTCAGAA 1319
|||
Qy 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512
|||
Db 1320 AATGTGAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAAATCTTATAGTGCAC 1379
|||
Qy 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
|||
Db 1380 TCGAGGCCAGCCGCAAAAAGAACCCCTCACAGTGTGTCTAATGGGTCTCCAGATTGTGATGTCT 1439
|||
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThrHisSer 552
|||
Db 1440 AAACCTACTAATCTCTCTCTGCTTCCACTTCCAGACTTTTGGCAGACACGTTCC 1499
|||
Qy 553 CysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLysPro 572
|||
Db 1500 TGCATATCTGAACATAGTTCAATCAATGTAATAATACTGCTGAAGAAATTAACCA 1559
|||
```

```
Qy 573 SerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
|||
Db 1560 AATCAGGTTGGAGGAGAACGTCACAAATCAAAAATGGAGTGTGGCTCAAGACCTGGATCA 1619
|||
Qy 593 ArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
|||
Db 1620 CAGGAAGG-TGCTATATAGTGGATGCTCTCTTGAGCTACACAAATTTCTCATGTAGAAAAAGAT 1678
|||
Qy 613 AspLeuPro 615
|||
Db 1679 GACTTACCT 1687
|||
```

Search completed: June 14, 2006, 01:46:27  
Job time : 8263 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 20:22:20 ; Search time 14775 Seconds  
(without alignments)  
11651.198 Million cell updates/sec

Title: US-10-644-084-1  
Perfect score: 2692  
Sequence: 1 cgtaggagagtgcacaggagc.....ttgattatggagtaatgggg 2692

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6366136 seqs, 31973710525 residues

Total number of hits satisfying chosen parameters: 12732272

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

1: gb\_env:\*  
2: gb\_pat:\*  
3: gb\_ph:\*  
4: gb\_pl:\*  
5: gb\_pr:\*  
6: gb\_ro:\*  
7: gb\_sta:\*  
8: gb\_sy:\*  
9: gb\_un:\*  
10: gb\_vi:\*  
11: gb\_ov:\*  
12: gb\_hcg:\*  
13: gb\_in:\*  
14: gb\_om:\*  
15: gb\_ba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2692	100.0	2692	6 AF532969	Mus muscu
2	2692	100.0	3424	6 AK173064	Mus muscu
3	2688.8	99.9	3425	6 BC021749	Mus muscu
4	2676	99.4	3410	6 BC031527	Mus muscu
5	2189.4	81.3	3195	6 AF532970	Rattus no
6	1810.2	67.2	2547	6 BC078687	Rattus no
7	1701.4	63.2	3220	5 BC033637	Homo sapi
8	1701.4	63.2	5835	2 CQ722034	Sequence
9	1699.8	63.1	5835	5 AB023140	Homo sapi
10	1512.2	56.2	3073	5 AY651262	Homo sapi
11	1492.6	55.4	2379	5 AY367055	Homo sapi
12	1290.6	47.9	2540	5 BC064389	Homo sapi
13	1260.4	46.8	2716	2 BD156617	Primer fo
14	1260.4	46.8	2716	2 AX877359	Sequence
15	1260.4	46.8	2716	5 AK001710	Homo sapi
16	1258.8	46.8	2272	2 AR718997	Sequence
17	1258.8	46.8	2272	2 AX747972	Sequence
18	1258.8	46.8	2272	5 AK093173	Homo sapi

19	944	35.1	165330	6	AC124987	Mus muscu
c 20	895.2	33.3	185929	12	AC017085	AC017085 Homo sapi
c 21	895.2	33.3	188840	5	AC046144	AC046144 Homo sapi
c 22	698.2	25.9	255776	12	AC142185	AC142185 Rattus no
c 23	698.2	25.9	306870	12	AC098557	AC098557 Rattus no
c 24	589	25.6	233457	12	AC106173	AC106173 Rattus no
c 25	560	20.8	826	2	BD148918	BD148918 Primer fo
c 26	560	20.8	826	2	AX868856	AX868856 Sequence
c 27	559.8	20.8	3892	11	BC070749	BC070749 Xenopus l
c 28	547.8	20.3	2594	11	BC060934	BC060934 Danio rer
c 29	530.8	19.7	3218	11	BC070587	BC070587 Xenopus l
c 30	421.8	15.7	781	11	EX934353	EX934353 Gallus ga
c 31	399.8	14.9	161874	5	AC114482	AC114482 Homo sapi
c 32	399.8	14.9	179844	5	AC009265	AC009265 Homo sapi
c 33	338.8	12.6	1503	2	BD158988	BD158988 Primer fo
c 34	338.8	12.6	1503	2	AX881380	AX881380 Sequence
c 35	338.8	12.6	1503	5	AK021515	AK021515 Homo sapi
c 36	334.2	12.4	563	2	BD150903	BD150903 Primer fo
c 37	334.2	12.4	563	2	AX870841	AX870841 Sequence
c 38	323	12.0	498	2	BD058793	BD058793 Sequence
c 39	312.2	11.6	674	11	CR353768	CR353768 Gallus ga
c 40	305.8	11.4	403	2	AX071986	AX071986 Sequence
c 41	297.2	11.0	428	2	BD026751	BD026751 Sequence
c 42	297.2	11.0	428	2	AR727012	AR727012 Sequence
c 43	297.2	11.0	428	2	AX887141	AX887141 Sequence
c 44	218.2	8.1	258641	12	AC127120	AC127120 Rattus no
c 45	212	7.9	884	2	BD148462	BD148462 Primer fo

#### ALIGNMENTS

RESULT 1  
AF532969  
LOCUS Mus musculus afadin- and alpha-actinin-binding protein ADIP mRNA, 2692 bp linear ROD 03-FEB-2003  
DEFINITION complete cds.  
ACCESSION AF532969  
VERSION AF532969.1 GI:27451838  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2692)  
AUTHORS Asada, M., Irie, K., Morimoto, K., Yamada, A., Ikeda, W., Takeuchi, M. and Takai, Y.  
TITLE ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized at Cell-Cell Adherens Junctions  
J. Biol. Chem. 278 (6), 4103-4111 (2003)  
JOURNAL 12446711  
PUBMED 2 (bases 1 to 2692)  
REFERENCE 2 (bases 1 to 2692)  
AUTHORS Asada, M., Irie, K. and Takai, Y.  
TITLE Direct Submision  
JOURNAL Submitted (26-JUL-2002) Department of Molecular Biology and Biochemistry, Osaka University Graduate School of Medicine/Faculty of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan

#### FEATURES

source  
1..2692  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
80..1927  
/codon\_start=1  
/product="afadin- and alpha-actinin-binding protein ADIP"  
/protein\_id="AA015015.1"  
/db\_xref="GI:27451839"  
/translation="MGDMVTVPVLCTENKNLSQYTSKTMSPSSLYSQVLCSSVP LSKNVHGVCFCTGENIQSISYLDQELTTFGPSPSYEKSKEAKENIVAVLNC MNELLVQRKNLLAQSVETQNKJGSDMDHLQSCYAKLKEQETSRRMIGLQRRR QLOCKNRSUHLKNEKDEVQKLQNIASRATQYNHDKRKREYNKLERLHQLVMN

KKDKNIAMDVLNVVVRADGKRGSWRTDKTEARNEDEMYKILLNDYEVROKQILMENAE  
LKKVLQOMKXEMISLLSPQKKPRERAEDGTGTVASDIEDDDSELSDRSVWGLSCDT  
VREQLTINSIRKQWILKSHVEKLDNQASKVHSEGLNEEDVTSRODHSEQETEKLELEIE  
RKEMIQAQOQLLOQOLATTCDDTTLSLLRDCYLLEBKERLKEWTLFKEQKNFERE  
RRSFTEAIRLGLERKAFEEBRASWVKQFLNMTNPFHQNSENVKLFSAFSGSDPDN  
LIVHSRPROKKLHVSANGVPACTSKLTKSPASPTSDFRQTHSCVSEHSSISVLNIT  
PEESKPSEVARESTDQKWSVQSRSSREGCYSGCSSAFRSHAGDRDLP"

ORIGIN

		Query Match	100.0%;	Score 2692;	DB 6;	Length 2692;				
		Best Local Similarity	100.0%;	Pred. No. 0;						
		Matches 2692;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;				
Qy	1	CGTAGGAGAGTCAGGAGCTGTTGTAAGCGTCGCAGCACTGAGCCGCCCTCTCAGGTAT	60							
Db	1	CGTAGGAGAGTCAGGAGCTGTTGTAAGCGTCGCAGCACTGAGCCGCCCTCTCAGGTAT	60							
Qy	61	CCTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120							
Db	61	CCTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120							
Qy	121	AGAAAAAATAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCGTCCAGTTTGTGA	180							
Db	121	AGAAAAAATAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCGTCCAGTTTGTGA	180							
Qy	181	CTCCAGCAAGTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGG	240							
Db	181	CTCCAGCAAGTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGG	240							
Qy	241	TGTCTTTCTGCACAGGAGAGAACATTGAACAAAGTATTTCTCTATCTTGATCAGGAGCTGAC	300							
Db	241	TGTCTTTCTGCACAGGAGAGAACATTGAACAAAGTATTTCTCTATCTTGATCAGGAGCTGAC	300							
Qy	301	CACCTTCGGGTTTCCTTCTGTTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAATT	360							
Db	301	CACCTTCGGGTTTCCTTCTGTTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAATT	360							
Qy	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGCTTCAGCGGAGAACCT	420							
Db	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGCTTCAGCGGAGAACCT	420							
Qy	421	GCTGGCCCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGCAGTGACATGGACCACT	480							
Db	421	GCTGGCCCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGCAGTGACATGGACCACT	480							
Qy	481	GCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTGGAACGTTCCAGCGCGGAGATGATCGG	540							
Db	481	GCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTGGAACGTTCCAGCGCGGAGATGATCGG	540							
Qy	541	GCTTCAAGAGAGACAGCGAGCTGCAGTCAAGAACAGGAGTTTGATCAGCTCTCTGAA	600							
Db	541	GCTTCAAGAGAGACAGCGAGCTGCAGTCAAGAACAGGAGTTTGATCAGCTCTCTGAA	600							
Qy	601	GAATGAGAAAGATGAGGTACAAAAATTAACAAAATATCATAGCCAGCGCGGCTACTCAGTA	660							
Db	601	GAATGAGAAAGATGAGGTACAAAAATTAACAAAATATCATAGCCAGCGCGGCTACTCAGTA	660							
Qy	661	TAATCATGATGTGAAGAGGAGAGCGTGAATATAATAGCTAAAGAGCGCCTGCATCA	720							
Db	661	TAATCATGATGTGAAGAGGAGGAGCGTGAATATAATAGCTAAAGAGCGCCTGCATCA	720							
Qy	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTAAATTAATGTGGGTG	780							
Db	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTAAATTAATGTGGGTG	780							
Qy	781	AGCTGATGGCAACGAGGCTCATGAGGACTGACAAAACAGAGCCAGGAATGAAGATGA	840							
Db	781	AGCTGATGGCAACGAGGCTCATGAGGACTGACAAAACAGAGCCAGGAATGAAGATGA	840							
Qy	841	GATGTACAAAATTCCTGTTGAATGATTAAGTACCGCCAGAGCAGATCTCTGATGGAGAA	900							
Db	841	GATGTACAAAATTCCTGTTGAATGATTAAGTACCGCCAGAGCAGATCTCTGATGGAGAA	900							

Qy	901	CGCGGAGCTCAAGAGGTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC	960
Db	901	CGCGGAGCTCAAGAGGTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCC	960
Qy	961	TCAGAAGAAGAACCCAGGGAAGAGCAGAGGACGGCAGAGGCACTGTGTCTATCTCCGA	1020
Db	961	TCAGAAGAAGAACCCAGGGAAGAGCAGAGGACGGCAGAGGCACTGTGTCTATCTCCGA	1020
Qy	1021	TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTTCTCTGTGAC	1080
Db	1021	TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTTCTCTGTGAC	1080
Qy	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGAAAACAGTCGAGAAATTTTGAAAAGTCATGT	1140
Db	1081	TGTGAGAGCAGCTGACAAAACAGCATCAGAAAACAGTCGAGAAATTTTGAAAAGTCATGT	1140
Qy	1141	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGGCTTTAATGAGGAGACGT	1200
Db	1141	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGGCTTTAATGAGGAGACGT	1200
Qy	1201	CATCTCAGCAAGACCATCAGCAAGACTGAGAACTGGAGCTGGAGATTGAGCGGTG	1260
Db	1201	CATCTCAGCAAGACCATCAGCAAGACTGAGAACTGGAGCTGGAGATTGAGCGGTG	1260
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCCACACCTGTGA	1320
Db	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGCCACACCTGTGA	1320
Qy	1321	TGATGACACCACTCACTGTTCGAGACTGTCTTACTTGTGGAAGAAAAGGAAACGCTTAA	1380
Db	1321	TGATGACACCACTCACTGTTCGAGACTGTCTTACTTGTGGAAGAAAAGGAAACGCTTAA	1380
Qy	1381	AGAAGAGTGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGAAAGCTT	1440
Db	1381	AGAAGAGTGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGGCGAAAGCTT	1440
Qy	1441	TACAGAGCTGCCATTCGATTGGGTTGGAGAGAAAGCGCTTTGAAGAAGAGCGAGCCAG	1500
Db	1441	TACAGAGCTGCCATTCGATTGGGTTGGAGAGAAAGCGCTTTGAAGAAGAGCGAGCCAG	1500
Qy	1501	CTGGTAAAGCAGCAGCTTTTAAACATGACGAACTTTTGACCAACAGAAATCTCAGAAAATGT	1560
Db	1501	CTGGTAAAGCAGCAGCTTTTAAACATGACGAACTTTTGACCAACAGAAATCTCAGAAAATGT	1560
Qy	1561	GAACCTTTTCAGTGCCCTCTCAGAGAGTCTGATCCAGACAACTTATAGTCCACTCACG	1620
Db	1561	GAACCTTTTCAGTGCCCTCTCAGAGAGTCTGATCCAGACAACTTATAGTCCACTCACG	1620
Qy	1621	GCCACGCAAAAGAGCTACACAGTGTGGCTTAATGGGTTGCCAGCTTGCACATCAAAACT	1680
Db	1621	GCCACGCAAAAGAGCTACACAGTGTGGCTTAATGGGTTGCCAGCTTGCACATCAAAACT	1680
Qy	1681	GACTAAATCTCTTCCTGCTCAGCTTCTGATCCAGAGTTCGCCAGACACATTCATGTGT	1740
Db	1681	GACTAAATCTCTTCCTGCTCAGCTTCTGATCCAGAGTTCGCCAGACACATTCATGTGT	1740
Qy	1741	GTCTGAACACAGTTCATCAGTGTGCTGAATATACTCTGAAAGAAAGTAAACCAAGTGA	1800
Db	1741	GTCTGAACACAGTTCATCAGTGTGCTGAATATACTCTGAAAGAAAGTAAACCAAGTGA	1800
Qy	1801	GGTTGCAAGAAAGACGAGTGTGAGCGTGTGAGCGTGTGAGCGGCGGCTCGCGGGA	1860
Db	1801	GGTTGCAAGAAAGACGAGTGTGAGCGTGTGAGCGTGTGAGCGGCGGCTCGCGGGA	1860
Qy	1861	GGGTGTCTACGGGATGTCTCGGCTTTCAGAGCGCTCAACGGGACCGAGATGACTT	1920
Db	1861	GGGTGTCTACGGGATGTCTCGGCTTTCAGAGCGCTCAACGGGACCGAGATGACTT	1920
Qy	1921	ACCTTAAATCTGGGGCTGAGTGTCTTCCAGATGTGCGCTAGAGAGTTGACACAGG	1980
Db	1921	ACCTTAAATCTGGGGCTGAGTGTCTTCCAGATGTGCGCTAGAGAGTTGACACAGG	1980
Qy	1981	GTGTAGCATAAAGTCAGTCTCTAACTTAAGATGCTCAGAGTTGTTTGTGGACTTCGC	2040





QY 421 GCTGCCAGGAGAGCGTGGAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCACCT 480  
DB 613 GCTGGCCAGAGAGCGTGGAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCACCT 672  
QY 481 CGAGAGCTGCTACGCGCAAACTTAAAGGAGCAGTGTGAAAACGTCAGGCGGGAGATGATCGG 540  
DB 673 CGAGAGCTGCTACGCGCAAACTTAAAGGAGCAGTGTGAAAACGTCAGGCGGGAGATGATCGG 732  
QY 541 GCTTCAAGAGAGACAGCGAGCTGCAGTGCAGNACAGGAGTTTGCATCAGCTCCTGAA 600  
DB 733 GCTTCAAGAGAGACAGCGAGCTGCAGTGCAGNACAGGAGTTTGCATCAGCTCCTGAA 792  
QY 601 GAAAGAGAAAGATGAGGTACAAAATATACAAAATATATACAGCAGCGGGCTACTCAGTA 660  
DB 793 GAATGAGAAAGATGAGGTACAAAATATACAAAATATATACAGCAGCGGGCTACTCAGTA 852  
QY 661 TAATCATGATGTGAAGAGGAAGGAGCGTGAATATAATTAAGCTAAAGGAGCGCCTGCATCA 720  
DB 853 TAATCATGATGTGAAGAGGAAGGAGCGTGAATATAATTAAGCTAAAGGAGCGCCTGCATCA 912  
QY 721 GCTCGTTATGAACAAGAGGATATAAACAATAGCCATGGATGTTTAAATATGTGGGTG 780  
DB 913 GCTCGTTATGAACAAGAGGATATAAACAATAGCCATGGATGTTTAAATATGTGGGTG 972  
QY 781 AGCTGATGCGCAAAACGAGGCTCATGGAGACTGACAAAACAGAAACAGGAAATGAAGATGA 840  
DB 973 AGCTGATGCGCAAAACGAGGCTCATGGAGACTGACAAAACAGAAACAGGAAATGAAGATGA 1032  
QY 841 GATGTACAAAATCTGTGTAATGATTAATGAGTACCGCCAGAGCAGATCCTGATGGAGAA 900  
DB 1033 GATGTACAAAATCTGTGTAATGATTAATGAGTACCGCCAGAAACAGATCCTGATGGAGAA 1092  
QY 901 CGCGAGCTGAAGAAGTCTCCAGCAGATGAAGAGAGAGATGATCTCTCCTGTCTCC 960  
DB 1093 CGCGAGCTGAAGAAGTCTCCAGCAGATGAAGAGAGAGATGATCTCTCCTGTCTCC 1152  
QY 961 TCAGAAAGAAAGCCAGGAAAGAGCAGAGAGAGCGCACAGGCACTGTGCTATCTCCGA 1020  
DB 1153 TCAGAAAGAAAGCCAGGAAAGAGCAGAGAGAGCGGCACAGGCACTGTGCTATCTCCGA 1212  
QY 1021 TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1080  
DB 1213 TATAGAAGATGACTCTGGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1272  
QY 1081 TGTGAGAGAGAGCTGACAAACAGCATCAGAAACAGTGGAGAAATTTGAAAAGTCATGT 1140  
DB 1273 TGTGAGAGAGAGCTGACAAACAGCATCAGAAACAGTGGAGAAATTTGAAAAGTCATGT 1332  
QY 1141 AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGCGCTTAATGAGGAGGACGT 1200  
DB 1333 AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGCGCTTAATGAGGAGGACGT 1392  
QY 1201 CATCTCAGACAAGACCATAGCAAGAGACTGAGAAAACCTGGAGCTGGAGATTTAGCGGTG 1260  
DB 1393 CATCTCAGACAAGACCATAGCAAGAGACTGAGAAAACCTGGAGCTGGAGATTTAGCGGTG 1452  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTA 1320  
DB 1453 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTA 1512  
QY 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAAGGAAACGCTTAA 1380  
DB 1513 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAAGGAAACGCTTAA 1572  
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAAGAGAGAAAGGCGGAAGCTT 1440  
DB 1573 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAAGAGAGAAAGGCGGAAGCTT 1632  
QY 1441 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTTCAAGAGAGCGAGCCAG 1500  
DB 1633 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTTGAAGAGAGCGAGCCAG 1692  
QY 1501 CTGGGTAAAGCAGCAGTGTGTTTAAACATGACGAACTTTTGACCACCAAGAACTCAGAAAAATGT 1560

DB 1693 CTGGGTAAAGCAGCAGTGTGTTTAAACATGACGAACCTTTGACCAACAGAACTCAGAAAAATGT 1752  
QY 1561 GAAACTTTTTCAGTGCCCTTTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAGC 1620  
DB 1753 GAAACTTTTTCAGTGCCCTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAGC 1812  
QY 1621 GCCACGCGCAAGAGAGCTACACAGTGTGGCTTAATGGGTGCCAGCTTGCACATCAAAACT 1680  
DB 1813 GCCACGCGCAAGAGAGCTACACAGTGTGGCTTAATGGGTGCCAGCTTGCACATCAAAACT 1872  
QY 1681 GACTTAAATCTCTCTCGCTCCTCAGCTTCTACTTTCAGACTTTCGCCAGACACATTCATGTGT 1740  
DB 1873 GACTTAAATCTCTCTCGCTCCTCAGCTTCTACTTTCAGACTTTCGCCAGACACATTCATGTGT 1932  
QY 1741 GTCTGAACACAGTGTCCATCAGTGTGCTGAATATAAATCTCTGAAAGAAAGTAAACCAAGTGA 1800  
DB 1933 GTCTGAACACAGTGTCCATCAGTGTGCTGAATATAAATCTCTGAAAGAAAGTAAACCAAGTGA 1992  
QY 1801 GGTTCGAAGAGAAAGCAGGATCAGAGTGGAGCGTGCAGTTCGAGGCCAGCTCGCCGGA 1860  
DB 1993 GGTTCGAAGAGAAAGCAGGATCAGAGTGGAGCGTGCAGTTCGAGGCCAGCTCGCCGGA 2052  
QY 1861 GGGGTGCTACAGCGGATGCTCTCGGCTTTCAGGAGCGCTTCAGGGGACCGAGATGACTT 1920  
DB 2053 GGGGTGCTACAGCGGATGCTCTCGGCTTTCAGGAGCGCTTCAGGGGACCGAGATGACTT 2112  
QY 1921 ACCTTAAATGTGGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGGAGTTGACACAGG 1980  
DB 2113 ACCTTAAATGTGGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGGAGTTGACACAGG 2172  
QY 1981 GTGTAGCATAAAGTCAAGTGTCTAACTTAAAGTGTCTCAGAGTGTGTTGTTGGACTTCGC 2040  
DB 2173 GTGTAGCATAAAGTCAAGTGTCTAACTTAAAGTGTCTCAGAGTGTGTTGTTGGACTTCGC 2232  
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAAAGGATGCAAAAGCTTTTGGTTGT 2100  
DB 2233 TGTCTTCCCCCAAGAGCTGAAATGCTAAGCTACTTAAAAAGGATGCAAAAGCTTTGGTTGT 2292  
QY 2101 GTGTAGTAAACAGAACCCCTGCTGTGACTGCGAGGAATGATGCGGCTTTGATGGAA 2160  
DB 2293 GTGTAGTAAACAGAACCCCTGCTGTGACTGCGAGGAATGATGCGGCTTTGATGGAA 2352  
QY 2161 ACAGAGCGCTGGAATGATTGCTCGCAGGTACCGAGAGAGACACTTTTAGGGACTGGT 2220  
DB 2353 ACAGAGCGCTGGAATGATTGCTCGCAGGTACCGAGAGAGACACTTTTAGGGACTGGT 2412  
QY 2221 TCCTGTAAACATTTAAATATTTCTGTCCTCAAGTGTGTTGGCAATTTGGAAGTGTAGCTTTTACT 2280  
DB 2413 TCCTGTAAACATTTAAATATTTCTGTCCTCAAGTGTGTTGGCAATTTGGAAGTGTAGCTTTTACT 2472  
QY 2281 TGAATGTATCTGTAGATTTTAAACAGCAGGTTCTATATTTATTTATGTTTAGTGTGAT 2340  
DB 2473 TGAATGTATCTGTAGATTTTAAACAGCAGGTTCTATATTTATTTATGTTTAGTGTGAT 2532  
QY 2341 TTTGGGATTTACCTCTTTTATATGTTTGTCTGTACATATAATACATGACTATGTATA 2400  
DB 2533 TTTGGGATTTACCTCTTTTATATGTTTGTCTGTACATATAATACATGACTATGTATA 2592  
QY 2401 GAGGCTTTTAAAGTTTTAAACCTTCACACCATGCTTGAGTATAGCATTTTCATGCAATTA 2460  
DB 2593 GAGGCTTTTAAAGTTTTAAACCTTCACACCATGCTTGAGTATAGCATTTTCATGCAATTA 2652  
QY 2461 AATGTTTTTTCAGTGGCATGTTTACAGAGTTTAGNACCACTGCCACATCAGCAGTTAAGA 2520  
DB 2653 AATGTTTTTTCAGTGGCATGTTTACAGAGTTTAGNACCACTGCCACATCAGCAGTTAAGA 2712  
QY 2521 CTTTATTTTTTAAAGCCATCTGGGCAATAAAAAATTTCAAGGCCCTTTCAATAGCTGAGTTTCAAG 2580  
DB 2713 CTTTATTTTTTAAAGCCATCTGGGCAATAAAAAATTTCAAGGCCCTTTCAATAGCTGAGTTTCAAG 2772  
QY 2581 ATAACTAGAACTACTAACTGTTTATTTTGAAGATTTTAAAGCAATTTTATTTTATTTAT 2640



Db 2773 ATAACCTAGAACTACTAACCTTACATTTTGGAGATTTTAAAGCATTTGATTTATTTAT 2832  
Qy 2641 ATATGTGAATGTTATTAATTTCTAAGAGGAATTTGATTTATGAGTAATGGGG 2692  
Db 2833 ATATGTGAATGTTATTAATTTCTAAGAGGAATTTGATTTATGAGTAATGGGG 2884

RESULT 3  
LOCUS BC021749 3425 bp mRNA linear ROD 18-JUL-2005  
DEFINITION Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,  
mRNA (CDNA clone MGC:25823 IMAGE:4165430), complete cds.  
ACCESSION BC021749  
VERSION BC021749.1 GI:18256805  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 3425)

REFERENCE  
AUTHORS Strausberg, R.B., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Mammalian Gene Collection Program Team  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 3425)

NIH MGC Project  
Direct Submission  
Submitted (18-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 30 Row: m Column: 16.  
Location/Qualifiers  
1. 3425  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"

FEATURES

source

gene  
/db\_xref="taxon:10090"  
/clone="MGC:25823 IMAGE:4165430"  
/tissue\_type="Salivary gland, 10 week old female mouse"  
/clone\_lib="NCI CGAP\_SG2"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 3425  
/gene="Ssx2ip"  
/note="synonym: Adip"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
240-2087  
/gene="Ssx2ip"  
/codon\_start=1  
/product="synovial sarcoma, X breakpoint 2 interacting  
protein"  
/protein\_id="AAH21749.1"  
/db\_xref="GI:18256806"  
/db\_xref="GeneID:99167"  
/db\_xref="MGI:2139150"  
/translation="MGDMVTVPVLCNTENKLSQYTSKMSPSLSYQQVLCSSVP  
LSKNVHGVGFTGENTIEQISYLDQELTTFGPSPSYEESKSKAEKELNITAVLNC  
MELLVLRQKNLLAESVETONKLGSDMDHLQSCYAKLKEQLETSRREMGLQVRDR  
OLOCKNSLHOLLKNEDEVOKLONIIASRATQYNHVKRERENKLERLHQLVMN  
KDXNIANDVLNYVRADGKRGSRWTKTEARNEDEMYKILLNDEYKQKQLMENAE  
LKKVLOQMKKEMISLLSPQKKPRERAEDGTGTVAISIEDSDSGLSDSVGLSCDT  
VREQTNSIRKQWILSKHVEKLDNQASKVHSEGLNEEDVISRODHEQETVKLEIE  
RCKEMIKAQOOLLOQALATCDDDTTSLRRCLLYLEERKEWTLFQOKKNPERE  
RRSTEAARILGLERKAFEEERASWVKQFLNMTNFHONSENKVLKFAFGSSDPDN  
LIVHSRPROKHLHSHVANGVPACTSKLTKSLPASPSDFQTHSCVSHSISVLNIT  
PEESKPSEVARESTDQKWSQVSRSSREGCYSGCSSAPRSRSHGDRDDLP"

ORIGIN

Query Match 99.9%; Score 2688.8; DB 6; Length 3425;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2690; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CGTAGGAGTGCAGGAGCTGTTGTAAGCGTCGACGACCTGAGCCGCTCTCTCAGGTAT 60  
Db 161 CGTAGGAGTGCAGGAGCTGTTGTAAGCGTCGACGACCTGAGCCGCTCTCTCAGGTAT 220  
Qy 61 CCTGCCTCGAACTTGTATGGGAGATTGATGACCTGTGACAGATCCAGTTCGTGTGATC 120  
Db 221 CCTGCCTCGAACTTGTATGGGAGATTGATGACCTGTGACAGATCCAGTTCGTGTGATC 280  
Qy 121 AGAAAAAATAATCTCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTGTA 180  
Db 281 AGAAAAAATAATCTCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTGTA 340  
Qy 181 CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTCGATGTTGTTTCGG 240  
Db 341 CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTCGATGTTGTTTCGG 400  
Qy 241 TGTCTTCTGCACAGGAGACATTTGAACAAGATTTTCTCTATCTTGATCAGGAGCTGAC 300  
Db 401 TGTCTTCTGCACAGGAGACATTTGAACAAGATTTTCTCTATCTTGATCAGGAGCTGAC 460  
Qy 301 CACCTTCGGGTTTCTCTTGTATGAAGATCCAAAGATGAAGAGGCAAGAGAGAAATT 360  
Db 461 CACCTTCGGGTTTCTCTTGTATGAAGATCCAAAGATGAAGAGGCAAGAGAGAAATT 520  
Qy 361 AAATATAGTCGCTTTCTGAACTGTATGAACAGAGCTGCTGCTGTTTACGCGGAAGAACCT 420  
Db 521 AAATATAGTCGCTTTCTGAACTGTATGAACAGAGCTGCTGCTGTTTACGCGGAAGAACCT 580  
Qy 421 GCTGCCCCAGGAGCGCTGGAGACACAGAACTTGAAGCTGGGCAATGACATGGACCACCT 480  
Db 581 GCTGCCCCAGGAGCGCTGGAGACACAGAACTTGAAGCTGGGCAATGACATGGACCACCT 640  
Qy 481 GCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTTGAAGACGTCGCCAGGGGAGATGATCGG 540  
Db 641 GCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTTGAAGACGTCGCCAGGGGAGATGATCGG 700

QY 541 GCTTCAAGAGAGACAGGCGAGCTGCAGTCAAGAACAGAGAGTTTGCATCAGCTCCTGAA 600  
DB 701 GCTTCAAGAGAGACAGGCGAGCTGCAGTCAAGAACAGAGAGTTTGCATCAGCTCCTGAA 760  
QY 601 GAATGAGAAAGATGAGGTACAAAATATACAAAATATCATAGCCAGCGGGCTACTCAGTA 660  
DB 761 GAATGAGAAAGATGAGGTACAAAATATACAAAATATCATAGCCAGCGGGCTACTCAGTA 820  
QY 661 TAATCATGATGTGAAGAGAAAGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 720  
DB 821 TAATCATGATGTGAAGAGAAAGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 880  
QY 721 GCTGTTATGAACAGAGAGTAATAACATAGCCATGATGTTTTAAATATATGCGGTG 780  
DB 881 GCTGTTATGAACAGAGAGTAATAACATAGCCATGATGTTTTAAATATATGCGGTG 940  
QY 781 AGCTGATGGCAAAAGAGCTCATGGAGACTGACAAAACAGAAAGCCAGGAATGAAGATGA 840  
DB 941 AGCTGATGGCAAAAGAGCTCATGGAGACTGACAAAACAGAAAGCCAGGAATGAAGATGA 1000  
QY 841 GATGTACAAAATCTGTTGAATGATATAGTACCGCCAGAGAGAGATCCTGATGGAGAA 900  
DB 1001 GATGTACAAAATCTGTTGAATGATATAGTACCGCCAGAGAGAGATCCTGATGGAGAA 1060  
QY 901 CGCGAGCTGAAGAGGTCTCCAGCAGATGAAGAGAGAGATGATCTCTCTGCTCTCC 960  
DB 1061 CGCGAGCTGAAGAGGTCTCCAGCAGATGAAGAGAGAGATGATCTCTCTGCTCTCC 1120  
QY 961 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGAGCGGCACAGGCACTGTGCTATCTCCGA 1020  
DB 1121 TCAGAAGAAGAGCCAGGGAAAGAGCAGAGAGCGGCACAGGCACTGTGCTATCTCCGA 1180  
QY 1021 TATAGAGATGATCTGCGGAACTAGCAGAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1080  
DB 1181 TATAGAGATGATCTGCGGAACTAGCAGAGAGACAGCGTGTGGGCGCTTCTCTGTGACAC 1240  
QY 1081 TGTGAGAGAGAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTGAAAAGTCATGT 1140  
DB 1241 TGTGAGAGAGAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTGAAAAGTCATGT 1300  
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT 1200  
DB 1301 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT 1360  
QY 1201 CATCTCAGCAAGACCAATGAGCAAGAGACGTGAGAAACTGAGAGCTGAGAGATTGAGCGGTG 1260  
DB 1361 CATCTCAGCAAGACCAATGAGCAAGAGACGTGAGAAACTGAGAGCTGAGAGATTGAGCGGTG 1420  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTGA 1320  
DB 1421 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTGA 1480  
QY 1321 TGATGACACCACTCACTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAAAGCCCTTAA 1380  
DB 1481 TGATGACACCACTCACTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAAAGCCCTTAA 1540  
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGCAAGCTT 1440  
DB 1541 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGCAAGCTT 1600  
QY 1441 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGCAGCAG 1500  
DB 1601 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGCAGCAG 1660  
QY 1501 CTGGGTAAAGCAGCAGTTTTTAAACATGACGAACCTTTGACCAACAGAACTCAGAAAAATGT 1560  
DB 1661 CTGGGTAAAGCAGCAGTTTTTAAACATGACGAACCTTTGACCAACAGAACTCAGAAAAATGT 1720  
QY 1561 GAAACTTTTTCAGTGCCTCTCAGGAAGTTCTGATCCAGACAACTTATAGTCCACTCACG 1620  
DB 1721 GAAACTTTTTCAGTGCCTCTCAGGAAGTTCTGATCCAGACAACTTATAGTCCACTCACG 1780  
QY 1621 GCCACGGCAAAAGAGCTACACAGTGTGGCTAAATGGGGTGCAGCTTGCACATCAAAACT 1680

DB 1781 GCCACGGCAAAAGAGCTACACAGTGTGGCTAATGGGTGCCAGCTTGCACATCAAAACT 1840  
QY 1681 GACTAAATCTCTTCCGCTCCACTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGT 1740  
DB 1841 GACTAAATCTCTTCCGCTCCACTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGT 1900  
QY 1741 GTCTGAAACAGTTCCATCAGTGTGCTGAATATAAATCCCTGAAAGAAAGTAAACCAAGTGA 1800  
DB 1901 GTCTGAAACAGTTCCATCAGTGTGCTGAATATAAATCCCTGAAAGAAAGTAAACCAAGTGA 1960  
QY 1801 GGTTCGAAAGAAAGACCGGATCAGAAAGTGAGCGTGCAGTGCAGAGCCAGCTCGCGGA 1860  
DB 1961 GGTTCGAAAGAAAGACCGGATCAGAAAGTGAGCGTGCAGTGCAGAGCCAGCTCGCGGA 2020  
QY 1861 GGGTGTCTACAGCGGATGCTCTCGGCTTCAGAGCGCTCACGGGGACCGAGATGACTT 1920  
DB 2021 GGGTGTCTACAGCGGATGCTCTCGGCTTCAGAGCGCTCACGGGGACCGAGATGACTT 2080  
QY 1921 ACCTTAAATGTGCGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGAGTTGACACAGG 1980  
DB 2081 ACCTTAAATGTGCGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGAGTTGACACAGG 2140  
QY 1981 GTGTAGCATAAAGTCAAGTGTCTTAAAGTGTCTCAGAGTTGTGTTTGGACTTCGC 2040  
DB 2141 GTGTAGCATAAAGTCAAGTGTCTTAAAGTGTCTCAGAGTTGTGTTTGGACTTCGC 2200  
QY 2041 TGTCTTCCCCAAAGAGCTGAAATGCTAAGCTACTTAAAGGATGCAAGCTTTGGTTGT 2100  
DB 2201 TGTCTTCCCCAAAGAGCTGAAATGCTAAGCTACTTAAAGGATGCAAGCTTTGGTTGT 2260  
QY 2101 GTCTAGTAAACAGAGCCCTGCTGTGACTGCAGGAATGATGCGCTTTGGATGCAA 2160  
DB 2261 GTGTAGTAAACAGAGCCCTGCTGTGACTGCAGGAATGATGCGCTTTGGATGCAA 2320  
QY 2161 ACAGAGCGCTGGAATGATTGCCCTCGCCAGGTACCGAGAGAGACACTTTTAGGACTTGGT 2220  
DB 2321 ACAGAGCGCTGGAATGATTGCCCTCGCCAGGTACCGAGAGAGACACTTTTAGGACTTGGT 2380  
QY 2221 TCCTGTAAACATTAATATTTTCGCTCCCAAGTGTGGTATGGAAGTGTAGCCTTTACT 2280  
DB 2381 TCCTGTAAACATTAATATTTTCGCTCCCAAGTGTGGTATGGAAGTGTAGCCTTTACT 2440  
QY 2281 TGAATGTATACGTGAGATTTTAAACAGCAGGTCTCTATATTTATTTAGTGTGAT 2340  
DB 2441 TGAATGTATACGTGAGATTTTAAACAGCAGGTCTCTATATTTATTTAGTGTGAT 2500  
QY 2341 TTTGGGATTAACCTCTTTTCATATGTTTTGTCTGTACATAAAATATACATGACTATGTAA 2400  
DB 2501 TTTGGGATTAACCTCTTTTCATATGTTTTGTCTGTACATAAAATATACATGACTATGTAA 2560  
QY 2401 GAGCTTTAAGGTTTAAAACTTCACACCATGCTTGAGTATAGCATTTTCATGCCAATTA 2460  
DB 2561 GAGCTTTAAGGTTTAAAACTTCACACCATGCTTGAGTATAGCATTTTCATGCCAATTA 2620  
QY 2461 AATGTTTTTCAGTGCAGTGTGTTTACAGAGGTTAGGACCACTGCACATGACAGTTAAGA 2520  
DB 2621 AATGTTTTTCAGTGCAGTGTGTTTACAGAGGTTAGGACCACTGCACATGACAGTTAAGA 2680  
QY 2521 CTTTATTTTTAAGCCATCTCGGCAATAAAAAATTTCAAGCCCCCTTCATAAGCTGAGTTTCA 2580  
DB 2681 CTTTATTTTTAAGCCATCTCGGCAATAAAAAATTTCAAGCCCCCTTCATAAGCTGAGTTTCA 2740  
QY 2581 ATAACTAGAACTACTAAGCTTACATTTTTTTCAGATTTTTTAAAGCAATTCATTTTATTTAT 2640  
DB 2741 ATAACTAGAACTACTAAGCTTACATTTTTTTCAGATTTTTTAAAGCAATTCATTTTATTTAT 2800  
QY 2641 ATATGGAATTTATATTTTCTAAGAGCAATATTTGATTTATGGAGTAATGGGG 2692  
DB 2801 ATATGGAATTTATATTTTCTAAGAGCAATATTTGATTTATGGAGTAATGGGG 2852

BC031527	BC031527	3410 bp	mRNA	linear	ROD 29-JUN-2004
LOCUS	Mus musculus synovial sarcoma, X breakpoint 2 interacting protein,				
DEFINITION	mRNA (CDNA clone MGC:28268 IMAGE:4008624), complete cds.				
ACCESSION	BC031527				
VERSION	BC031527.1	GI:21594536			
KEYWORDS	MGC.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 3410)				
	Strausberg, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Klausner, R.D., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywicki, M.I., Skalek, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 3410)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA				
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>				
COMMENT	Contact: MGC help desk Email: <a href="mailto:c9apbs-remail.nih.gov">c9apbs-remail.nih.gov</a> Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a> Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.				
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 36 Row: k Column: 10 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 20270272. Location/Qualifiers 1. .3410 /organism="Mus musculus" /mol_type="mRNA" /strain="Czech II" /db_xref="taxon:10090" /clone="MGC:28268 IMAGE:4008624" /tissue type="Mammary tumor metastatized to lung. MMTV-LTR/Mnt1 model. Expression driven by an MMTV-LTR enhancer." /clone_lib="NCI_CGAP_Lu30"				
source					
gene	/lab_host="DH10B" /note="Vector: pCMV-SPORT6" 1. .3410 /gene="Sex2ip" /note="synonym: Adip" /db_xref="GeneID:99167" /db_xref="MGI:2139150" 248. .2095 /gene="Sex2ip" /codon_start=1 /product="synovial sarcoma, X breakpoint 2 interacting protein" protein				
CDS					
	Query Match 99.4%; Score 2676; DB 6; Length 3410; Best Local Similarity 99.6%; Pred. No. 0; Matches 2682; Conservative 0; Mismatches 10; Indels 0; Gaps 0;				
	QY 1 CGTAGGAGGTGACAGGAGCTGTGTAAGCGTGCAGCACTGAGCGCGCTCTCTCAGGTAT 60 DB 169 CGTAGGAGGTGACAGGAGCTGTGTAAGCGTGCAGCACTGAGCGCGCTCTCTCAGGTAT 228 QY 61 CTGCTCTGGAACTTGTCTATGGAGATTGATGATCTGTGACAGATCAGATTCCTGTGTAC 120 DB 229 CTTGCTCTGGAACTTGTCTATGGAGATTGATGATCTGTGACAGATCAGATTCCTGTGTAC 288 QY 121 AGAAACAAAATCTCTCAATATACCTCAGAACAAAGATGCTCCGTCAGTTGTA 180 DB 289 AGAAACAAAATCTCTCAATATACCTCAGAACAAAGATGCTCCGTCAGTTGTA 348 QY 181 CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTCAGTGGTTCGG 240 DB 349 CTCCAGCAAGTCTGTGCTCTTCTCAGTACCTTTATCCAAAAACGTCAGTGGTTCGG 408 QY 241 TGTCTTCTGCACAGGAGAGAAACATTGAAACAAAGTATTTCTTATCTTGTATCAGGAGCTGAC 300 DB 409 TGTCTTCTGCACAGGAGAGAAACATTGAAACAAAGTATTTCTTATCTTGTATCAGGAGCTGAC 468 QY 301 CACCTTCGGGTTTCTTCTTCTTGTATGAGATCCAAAAGTAAAGAGGCAAGAGAGAAAT 360 DB 469 CACCTTCGGGTTTCTTCTTCTTGTATGAGAAATCCAAAAGTAAAGAGGCAAGAGAGAAAT 528 QY 361 AAATATAGTCCCTGTTCTTGAACCTGTATGAACGAGCTGCTCGTCTTACGCGAAGAACT 420 DB 529 AAATATAGTCCCTGTTCTTGAACCTGTATGAACGAGCTGCTCGTCTTACGCGAAGAACT 588 QY 421 GCTGCCCGCAGGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGCATGGACCACT 480 DB 589 GCTGCCCGCAGGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGCATGGACCACT 648 QY 481 GCAGAGCTGTACCCCAACTTAAGGAGCAGTTGGAACCTGCCAGCGGGAGATGATCGG 540 DB 649 GCAGAGCTGTACCCCAACTTAAGGAGCAGTTGGAACCTGCCAGCGGGAGATGATCGG 708 QY 541 GCTTCAAGAGAGACAGGAGCTGTCAGTGCACAAACAGAGAGTTTGCATCAGCTCCCTGAA 600 DB 709 GCTTCAAGAGAGACAGGAGCTGTCAGTGCACAAACAGAGAGTTTGCATCAGCTCCCTGAA 768				

QY 601 GAATGAGAAAGATGAGGTACAAAATACAAAATATCATAGCCAGCGGCTACTCAGTA 660  
DB |||||  
QY 769 GAATGAGAAAGATGAGGTACAAAATACAAAATATCATAGCCAGCGGCTACTCAGTA 828  
DB |||||  
QY 661 TAATCATGATGTGAAGAGGAAGGCGGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA 720  
DB |||||  
QY 829 TAATCATGATGTGAAGAGGAAGGCGGTGAATATAATAAGCTAAAGGAGCGCCTGCATCA 888  
DB |||||  
QY 721 GCTGTTATGAACAAGAGGATAAACAATAGCCATGATGTTTAAATATATGTTGGTGC 780  
DB |||||  
QY 889 GCTGTTATGAACAAGAGGATAAACAATAGCCATGATGTTTAAATATATGTTGGTGC 948  
DB |||||  
QY 781 AGCTGATGCGCAACGAGGCTCATGAGGACTGACAAAACAGAACGACGAGGATGAAGATGA 840  
DB |||||  
QY 949 AGCTGATGCGCAACGAGGCTCATGAGGACTGACAAAACAGAACGACGAGGATGAAGATGA 1008  
DB |||||  
QY 841 GATGTACAAAATCTCTGTTGAATGATATGAGTACCGCCAGAACGAGATCTCTGATGGAGAA 900  
DB |||||  
QY 1009 GATGTACAAAATCTCTGTTGAATGATATGAGTACCGCCAGAACGAGATCTCTGATGGAGAA 1068  
DB |||||  
QY 901 CGCGAGCTGAAGAGTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTGTTCTCC 960  
DB |||||  
QY 1069 CGCGAGCTGAAGAGTCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTGTTCTCC 1128  
DB |||||  
QY 961 TCAGAAGAAGAGCCAGGGAAGAGCAGAGAGCGCACAGGCACTGTTGCTATCTCCGA 1020  
DB |||||  
QY 1129 TCAGAAGAAGAGCCAGGGAAGAGCAGAGAGCGCACAGGCACTGTTGCTATCTCCGA 1188  
DB |||||  
QY 1021 TATAGAAGATGACTCTGGGGAACTGAGCAGAGACAGCGTGTGGGGCCCTTCTCTGTGACAC 1080  
DB |||||  
QY 1189 TATAGAAGATGACTCTGGGGAACTGAGCAGAGACAGCGTGTGGGGCCCTTCTCTGTGACAC 1248  
DB |||||  
QY 1081 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGTCAATGT 1140  
DB |||||  
QY 1249 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGTCAATGT 1308  
DB |||||  
QY 1141 AGAAAACCTCATAAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGACGT 1200  
DB |||||  
QY 1309 AGAAAACCTCATAAACCAAGCTTCGAAGGTGCATCTCAGAGGGCCCTTAATGAGGAGGACGT 1368  
DB |||||  
QY 1201 CATCTCAGCACAAGACCATGAGCAAGAGACTTGAGAAACTGGAGCTGGAGATTTGAGCGGTG 1260  
DB |||||  
QY 1369 CATCTCAGCACAAGACCATGAGCAAGAGACTTGAGAAACTGGAGCTGGAGATTTGAGCGGTG 1428  
DB |||||  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACCACTGTGA 1320  
DB |||||  
QY 1429 TAAAGAAATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACCACTGTGA 1488  
DB |||||  
QY 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAACGCTTAA 1380  
DB |||||  
QY 1489 TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAGGAACGCTTAA 1548  
DB |||||  
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGCAAGCTT 1440  
DB |||||  
QY 1549 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGCAAGCTT 1608  
DB |||||  
QY 1441 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGCCAG 1500  
DB |||||  
QY 1609 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGCCAG 1668  
DB |||||  
QY 1501 CTGGTAAAGCAGCAGTGTATTAACATGACGAACCTTTGACCACAGCAACTCAGAAAAATGT 1560  
DB |||||  
QY 1669 CTGGTAAAGCAGCAGTGTATTAACATGACGAACCTTTGACCACAGCAACTCAGAAAAATGT 1728  
DB |||||  
QY 1561 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCACG 1620  
DB |||||  
QY 1729 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCACG 1788  
DB |||||  
QY 1621 GCCAGGCAAAAGAGCTTACAGTGTGGCTTAATGGGGTGCAGCTTGACATCAAACT 1680  
DB |||||  
QY 1789 GCCAGGCAAAAGAGCTTACAGTGTGGCTTAATGGGGTGCAGCTTGACATCAAACT 1848  
DB |||||  
QY 1681 GACTAAATCTCTCTCTGCTCACCTTCTACTTCAGACTTTCCGCCAGACACACATTCATGTGT 1740  
DB |||||

DB |||||  
QY 1849 GACTAAATCTCTCTGCTCACTCTCTACTCTCAGACTTTCCGCCAGACACATTCATGTGT 1908  
DB |||||  
QY 1741 GTCTGAACACAGTTCATAGTGTGCTGAATATAAATCTCTGAAGAAAAGTAAACCAAGTGA 1800  
DB |||||  
QY 1909 GTCTGAACACAGTTCATAGTGTGCTGAATATAAATCTCTGAAGAAAAGTAAACCAAGTGA 1968  
DB |||||  
QY 1801 GGTTCGAAGAAAAGACACGATCAGAAAGTGCAGGCGTGCAGTCCAGAGGCCAGCTCGCCGGA 1860  
DB |||||  
QY 1969 GGTTCGAAGAAAAGACACGATCAGAAAGTGCAGGCGTGCAGTCCAGAGGCCAGCTCGCCGGA 2028  
DB |||||  
QY 1861 GGGTGTCTACAGCGGATGCTCTCGGCTTTCAGGAGCGCTCACGGGGAACCGAGATGACTT 1920  
DB |||||  
QY 2029 GGGTGTCTACAGTGGATGTTCTCGGCTTTCAGGAGCGCTCACGGGGAACCGAGATGACTT 2088  
DB |||||  
QY 1921 ACCTTAAATGTGGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGGAGTTGACACAGG 1980  
DB |||||  
QY 2089 ACCTTAAATGTGGGGCTGCAGTGTCTCCAGATGTGCGCTAGAGGAGTTGACACAGG 2148  
DB |||||  
QY 1981 GTGTAGCATAAAGTCAAGTCTGCTAACTTAAAGATGCTCAGAGTGTGTTTGGACTTTCGC 2040  
DB |||||  
QY 2149 GTGTAGCATAAAGTCAAGTCTGCTTAACTTAAAGATGCTCAGAGTGTGTTTGGACTTTCGC 2208  
DB |||||  
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTTAAGCTACTTAAAAAGGATGCAAAAGCTTTGGTGT 2100  
DB |||||  
QY 2209 TGTCTTCCCCCAAGAGCTGAAATGCTTAAGCTACTTAAAAAGGATGCAAAAGCTTTGGTGT 2268  
DB |||||  
QY 2101 GTGTAGTAAACAGAGCCCTGGCTCTGTGACTGTCAGGAATGCATGCGGTTTGATGGAA 2160  
DB |||||  
QY 2269 GTGTAGTAAACAGAGCCCTGGCTCTGTGCTGTCAGGAATGCATGCGGTTTGATGGAA 2328  
DB |||||  
QY 2161 ACAGAGCGCTGGAAATGATTGCTCCGACAGTACCGAGAGAGAGCACTTTTAGGAGACTGGT 2220  
DB |||||  
QY 2329 ACAGAGCGCTGGAAATGATTGCTCCGACAGTACCGAGAGAGAGCACTTTTAGGAGACTGGT 2388  
DB |||||  
QY 2221 TCCTGTAAACATTTAAATATTTTCGTCCTCCCAAGTGTGGTTGGCAATTTGGAAGTGTAGCCCTTACT 2280  
DB |||||  
QY 2389 TCCTGTAAACATTTAAATATTTTCGTCCTCCCAAGTGTGGTTGGCAATTTGGAAGTGTAGCCCTTACT 2448  
DB |||||  
QY 2281 TGAATGTATACTGTAGATTTTAAACAAAGCAGGTTCTATATTATTTATTTAGTGTGAT 2340  
DB |||||  
QY 2449 TGAATGTATACTGTAGATTTTAAACAAAGCAGGTTCTATATTATTTATTTAGTGTGAT 2508  
DB |||||  
QY 2341 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATACATACATGACTATGTAA 2400  
DB |||||  
QY 2509 TTTGGGATTAACCTCTTTTCATATGTTTGTCTGTACATAAAATACATGACTATGTAA 2568  
DB |||||  
QY 2401 GAGGCTTTAAGGTTTAAAAAATTTTCAACCATGCTTGAGTATAGCATTTTCATGCCAATTA 2460  
DB |||||  
QY 2569 GAGGCTTTAAGGTTTAAAAAATTTTCAACCATGCTTGAGTATAGCATTTTCATGCCAATTA 2628  
DB |||||  
QY 2461 AATGTTTTTCAGTGGCATGGTTTACAGAGGTTAGGACCACTGCCACATGACAGCTTAAGA 2520  
DB |||||  
QY 2629 AATGTTTTTCAGTGGCATGGTTTACAGAGGTTAGGACCACTGCCACATGACAGCTTAAGA 2688  
DB |||||  
QY 2521 CTTTATTTTTAAGCCATCTCGGCAATAAAAAATTTCAAAAGCCCTTTCATAAGCTGAGTTCAG 2580  
DB |||||  
QY 2689 CTTTATTTTTAAGCCATCTCGGCAATAAAAAATTTCAAAAGCCCTTTCATAAGCTGAGTTCAG 2748  
DB |||||  
QY 2581 ATAACTAGAACTACTAAACGTTTACATTTTTCAGATTTTAAAGCAATTTGATTTTATTTAT 2640  
DB |||||  
QY 2749 ATAACTAGAACTACTAAACGTTTACATTTTTCAGATTTTAAAGCAATTTGATTTTATTTAT 2808  
DB |||||  
QY 2641 ATATGCAATGTTATTAATTTTAAAGAGGAATATTGATTATGGAGTAATGGGG 2692  
DB |||||  
QY 2809 ATATGCAATGTTATTAATTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2860  
DB |||||

RESULT 5  
AF532970  
LOCUS  
DEFINITION Rattus norvegicus afadin- and alpha-actinin-binding protein ADIP  
mRNA, complete cds.

AF532970 3195 bp mRNA linear ROD 03-FEB-2003  
Rattus norvegicus afadin- and alpha-actinin-binding protein ADIP  
mRNA, complete cds.

ACCESSION	AF532970	Qy	362	AATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGCTGCTTACGCGGAAGAACCTG	421
VERSION	AF532970.1	Db	361	AGTATAGTTGCTCTTTCTGAACCTGCATGAATGAGCTGCTTGTGTTCTCAGCGAAGAACCTC	420
KEYWORDS					
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 3195)				
AUTHORS	Asada, M., Irie, K., Morimoto, K., Yamada, A., Ikeda, W., Takeuchi, M. and Takai, Y.				
TITLE	ADIP, a Novel Afadin- and alpha-Actinin-Binding Protein Localized at Cell-Cell Adherens Junctions				
JOURNAL	J. Biol. Chem. 278 (6), 4103-4111 (2003)				
PUBMED	12446711				
REFERENCE	2 (bases 1 to 3195)				
AUTHORS	Asada, M., Irie, K. and Takai, Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-JUL-2002) Department of Molecular Biology and Biochemistry, Osaka University Graduate School of Medicine/Faculty of Medicine, Yamada-oka 2-2, Suita 565-0871, Japan				
FEATURES	Location/Qualifiers				
source	1..3195				
	/organism="Rattus norvegicus"				
	/mol_type="mRNA"				
	/strain="Sprague-Dawley"				
	/db_xref="taxon:10116"				
	79..1920				
	/codon_start=1				
	/product="afadin- and alpha-actinin-binding protein ADIP"				
	/protein_id="AAO15016.1"				
	/db_xref="GI:27451841"				
CDS	/translation="MGDMWTVDVPLCTENKNSLYTSQYSETKMSPLSSLYQQVLSATP LSKNVHVSFAFCTGENTEQSYLDQELTTFPSPSLYEEKSKEARELSIVALLNC NNEILLVQRNKLQAESVETQNLKGLSDHMLQSCYAKLKEQLSEARREMSLOERDR QLOQNRNLHOLLNKEVEVKQKONI IASRATQYNDVVRKEREYKILKRLHOLVNR KKKNIAMDVLNIVGRVDGKSRWRTDKTEARNDVEMVKILLNDYEVQKQILLNENAE LKXVILQMKKEMISLLSPKKPRERAEDSTGVVISDVEDDAGELSRDGSWSLSCDT VREQLTNSIRKQWLKSLVEKLNQASVHSEGFHEEDVISRQDHQEOETKEJLEIE RCKEMIKAQOQLQQLATACDDDTLSLRDCYLLKEKERLKEWSLFEKQKNFERE RRSFTEAIRLGLRKAPEERASVWQOFLNMTTFDHNSKNVLSKLSFSAFGSGSDPON LIVHPRPRKPHSVANGVPACTSKLAKSLTPSPDFCPSRSCVSEHSFVSALVTPTPE ETKEVNGRESTDQKSVSVSRPSSREGCYGCGSSAYTSSHVERDDLPL"				
ORIGIN					
Query Match	81.3%; Score 2189.4; DB 6; Length 3195;				
Best Local Similarity	90.1%; Pred. No. 0;				
Matches	2427; Conservative 0; Mismatches 247; Indels 20; Gaps 7;				
Qy	2	GTAGGAGAGTACAGGAGCTGTTGTAAGCGTGCAGACACTGAGCGGCTCTCCTCAGGTATC	61		
Db	1	GTAGGAGAGTGACAGGAGCTGTTGTGCAATGCCCGACACTGAACTGCGCTTCTCAGGGACC	60		
Qy	62	CTGGCTCTGGAACTTGCTATGGAGATGGATGACTGTGACAGATCCAGTCTCTGTGTACA	121		
Db	61	CTGGCTCTGGAGCTGGCTATGGAGATGGATGACTGTGACAGATCCAGTCTCTGTGTACA	120		
Qy	122	GAATAACAAAATCTCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCAGGTTTGTAC	181		
Db	121	GAATAACAAAATCTCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCAGGTTTGTAC	180		
Qy	182	TCCAGGAGTTCGTGCTCTTCAGTACCTTTATCCAAAACCGTGATGGTGTGTTTCGGT	241		
Db	181	TCCAGGAGTTCGTGCTCTTCAGTACCTTTATCCAAAACCGTGATGGTGTGTTTCAGT	240		
Qy	242	GTCTTCTCCAGGAGAGAACATTGAACAAAGTATTTCTTCTATTCAGGAGCTGACC	301		
Db	241	GCCTTCTCCAGGAGAGAACATGAACAGAGTATTTCTGATCTTTGATCAGGAACTGACT	300		
Qy	302	ACCTTCGGGTTTCCTCTCTGTTGATGAAGATCCAAAAGTAAAGGCGCAAGAGAGAAATTA	361		
Db	301	ACCTTCGGTTCCTCTCTGTTGATGAAGATCCAAAAGTAAAGGCGCAAGAGAGTGA	360		

D	b	1441	ACAGAACTGCTATTTCGCTTTGGGTTCGAGAGAAAAGCGCTTTCGAGGAAGCAGCGCCAGC	1500
Q	y	1502	TGGTAAAGCAGCAGTGTTTTTAACATGACCAACTTTGCACCACCAGAACACTCAGAAAAATGTG	1561
D	b	1501	TGGGTGAAGCAGCAGTGTTTTTAACATGACCACTTTGATCACAGAACTCAGAAAAATGTG	1560
Q	y	1562	AACCTTTTCAGTGCCCTCTCAGGAAGTTCATGCCAGACAATCTTATAGTCCAATCAGCG	1621
D	b	1561	AACTTTTCAGTGCCCTTTTCAGGAAGTTCATGCCAGACAATCTTATAGTCCAATCAGCG	1620
Q	y	1622	CCACGGCAAGAAGCTACACAGGTGGCTTAATGGGGTGCACGCTTGCACATCAAAACTG	1681
D	b	1621	CCACGGCAAGAAGCCACACAGGTGCGCTAATGGGGTGCACGCTTGCACATCAAACTG	1680
Q	y	1682	ACTAAATCTCTTCTCGCTCACCTTCTACCTTCAGACTTTCGCCCAGACATTCATGTGTG	1741
D	b	1681	GCTAAGTCTCTTCCGACCTCAC-----CTTCAGACTTTCGCCCGTCTCGCTCATGTGTG	1734
Q	y	1742	TCTGAACACAGTTCCTCAGTGTGCTGAATAAATCTCTGAAGAAAGTAGTAAACCAAGTGAG	1801
D	b	1735	TCTGAGCACAGTCCCCTCAGTGGCTGACTGTGACTCTGGAAGAACCAACCGAATGAG	1794
Q	y	1802	GTTSCAAGAGAAGCACCGGATCAGAAGTGAGCGTGCAAGCGCCAGCTCGCGGGAG	1861
D	b	1795	GTTGGAAGAGAAGTACGGACCAAGAAGTGAGCGTGGTGTCCAGACCAGCTCCCGGGAG	1854
Q	y	1862	GGGTGCTACAGCGGATGCTCTCTGGCCCTTCAGAGCGCTCACGGGACCGAGATGACTTA	1921
D	b	1855	GGTTGCTACGGTGGATGCTCTCTGGCCCTCACAAAGCTCCCACGTGAACGAGATGACTTA	1914
Q	y	1922	CCTTAAATGTGCGGGCTGCAGTGTCTGTCACAGATGCGCTAGAGAGGTGACACAGGG	1981
D	b	1915	CCATAGGTGCTCGGGCTGCAAGCGCTGT--CCTGGAGTGCATGAGAGGAATTGACACGGG	1972
Q	y	1982	TGTAGCATAAAGTCAGTCGTCTAACTTAAGATGCTCAGAGTGTGTTGTTTGGACTTCGCT	2041
D	b	1973	TGTAGCATAAAGTCAGCCATCTACCGTAAGATG-TCGGAGTATTATTTGTTGGACTTCCCCA	2031
Q	y	2042	GTCCTTCCCCAAGAGCTGAAATGCTTAAGCTACTTAAAGAGTGCAAGCTTTGGTTGTG	2101
D	b	2032	GTCCTTCCCCAAGAGCTGAAACGCT-----TTAGAGGATCGAAAGCTTTGGCTGTG	2084
Q	y	2102	TGTTAGTAAACAGAGCCCTCGGCTCTGACTGCAGGAATGCATGGCGTTTGCATGGAAA	2161
D	b	2085	TGTTAGTAAACAGAGCCCTCGGCTCTGCTGAGTAAAGGAATGTATGTTGTTGGTGGAAA	2144
Q	y	2162	CAGAAGCGCTGGAAATGATTCCTC-GCCAGGTACCGAGAAAGACACCTTTTAGGGACTGGT	2220
D	b	2145	CAAAAGCACAGGAATGATTTCTCTCCGGGTACTGAGAATAGCACCTTTTAGGGACTGAT	2204
Q	y	2221	TCCTGTAACACTTAATATTTTCGTCGCCCAAGTGTGGTGGCAATCGGAAGGTAGCGTTTACT	2280
D	b	2205	TCCTGTAACACTTAATATTTTGTGCCCAAGTATGSGTTGGCAATCGGAAGTTAGTCTTTACT	2264
Q	y	2281	TGAATGTATACGTAGATTTTTTAACAAGCAGGTTCTATATTTATTTATGTTTAAAGTGTGAT	2340
D	b	2265	TGAATGTACCTGTAGATTTTTTAACAAGCA-GTTCCTATATTTATTTAGTTTAAAGTGTGAT	2323
Q	y	2341	TTTGGGATTACCTCTTTTCATATGTTTTGTGTCTGTACATAAATATACATCACTATGTTAA	2400
D	b	2324	TTTGGGATTACCTCTTTTCATATGTTTTCTGCGCTGTACATAAATATACATCACTATGTTAA	2383
Q	y	2401	GAGGCTTTAAGGTTTTAAAACTTCACACCATGCTTGGATGTAGCATTTTCATGCCAATTAA	2460
D	b	2384	GAGGCTTTAAGGTTTTAAAAATTTTCACACCATGCTCGAGNATAGCATTTTCATGCCAATTAA	2443
Q	y	2461	AATGTTTTTCAGTGGCATGGTGTTTTACAGA--GGTTAGGACCACTGCCACATGACAGTTAA	2518
D	b	2444	AATGTTTTTCAGTGGCATGGTGTTTTACAGATGTGTTTAGGACCACTGCCACATGACAGTTAA	2503
Q	y	2519	GACTTTATTTTTTAAGCCATCTGGGCAATAAAAAATTTCAAAGCCCTTCATTAAGCTGAGTTC	2578

Db	2504	GA	TTTTTATTTTAAAGCCATTTGGGCAATAAAAAATTCAAAGCCACTTCATTAAGCTAAGTTC	2563	
Qy	2579	AG	ATAACTAGAACTACTAAACGTTACATTTTGGAGATTTTAAAGCAATGTTATTTATTTT	2638	
Db	2584	AG	ATAGCTAAACTACTAACATTTTGGAGATTTTAAAGCAATTTATTTATTTT	2623	
Qy	2639	AT	ATATGTGGAATCTTATAATTTCTTAAGAGGAATATTGATTTATCGAGTAATGGGG	2692	
Db	2624	AT	ATATGTGACTCTTATAATTTCTTAAGAGGAATGTGGAATTATGAAGCAATGGGG	2677	
RESULT 6	BC078687	2547 bp	mRNA	linear	ROD 09-DEC-2005
LOCUS	Rattus norvegicus synovial sarcoma, X breakpoint 2 interacting protein, mRNA (cDNA clone MGC:93056 IMAGE:7115887), complete cds.				
ACCESSION	BC078687				
VERSION	GI:50926757				
KEYWORDS	MGC.				
SOURCE	Rattus norvegicus (Norway rat)				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 2547)				
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, B., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loguercio, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Munz, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzyzanski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.				
CONSRMT	Mammalian Gene Collection Program Team				
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)				
PUBMED	12477932				
REFERENCE	2 (bases 1 to 2547)				
AUTHORS	NTH MGC Project				
CONSRMT	Direct Submission				
TITLE	Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA				
JOURNAL	NTH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>				
REMARK	Contact: MGC help desk				
COMMENT	Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a>				
	Tissue Procurement: Howard Jacobs				
	cDNA Library Preparation: Express Genomics				
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305-0505				
	Web site: <a href="http://www-shgc.stanford.edu">http://www-shgc.stanford.edu</a>				
	Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a>				
	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>				
	Series: IRAC Plate: 163 Row: b Column: 1				
	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31342223				
FEATURES	Location/Qualifiers				

source	1. .2547 /organism="Rattus norvegicus" /mol_type="mRNA" /db_xref="taxon:10116" /clone="MGC:93056 IMAGE:7115887" /tissue_type="Testis, rat (Brown Norway)" /clone_lib="NIH MGC_237" /lab_host="DH10B" /note="Vector: pExpress1" 1. .2547 /gene="Sex2ip" /note="synonym: ADIP" /db_xref="GeneID:308023" /db_xref="RGD:708490" 366. .2207 /gene="Sex2ip" /codon_start=1 /product="synovial sarcoma, X breakpoint 2 interacting protein" /protein_id="AAH78687.1" /db_xref="GI:50926758" /db_xref="GeneID:308023" /db_xref="RGD:708490" /translation="MGDMWTVTDPLCTENKNLSOYTSETKMSPSLSYQOVLCSATP LSKNVGVFSAPCTGENTIEQSTSYLDDELTTGSPSLYEESKREKRELSIVALLNC MNELLVLORKNLLAQESVETQNLKSGMDHLQSCYAKLKEQLERSRREMISLQERDR QLQCKRNLHOLLKNEKEVEVKLQNIIASRATQYNHDKREKREYNLKERLHOLVWN RKDNAMVLYNGYVRDQKGRWRTDTEARNEDEMYKILLNDYERQKQILLSEAE LKVLQKQMKMISLLSPQKKPRAEADSTGVVIVSDVEDDAGELSDRGVSLSCDT VREOLTNSIRKQWRILKSHVEKLDNQASKVHSEGHREDVLSRODHEOTEKLELEIE RCKMIAQQOOLLQOLATACDDPTTSLRDCYLLEKEXLERLKEWSLFEQKNPFERE RRSTFEAIRGLERKATFEERASVWQQLNMVITDHQNSENVKLFSAFGSSDPDN LIVHPRPQKQKSVANVGPACTSKLAKSLPTSPDFCSRCSSEHSPVSALTVTPE ETKPEVNGRESTDQKSVSVSRPSREGCYGCGSSAYTSSHVERDLP"
gene	
CDS	
Db	767 GCAGAGCTGCTACGCCAACTTAAGGAAACAGTTGGAGGCTCCAGGCGAGAGATGATCAG 826
Qy	541 GCTTCAAGAGAGAGACAGCAGCTCAGTGCAAGAAACAGGAGTTTTCATCAGCTCCTGAA 600
Db	827 CTTTCAGAGAGAGACAGCAGCTACAGTGCAAAAAACAGGAATTTGCATCAGCTCCTGAA 886
Qy	601 GAATGAGAAAGATGAGTACAAAAATACAAATATCATAGCCAGCCGGCTACTCAGTA 660
Db	887 AAACGAGAAAGAGAGTACAAAAATACAAAAATATCATAGCCAGTCGGGCTACTCAGTA 946
Qy	661 TAAATCATGCTGAAGAGAGAGGAGCGTGAATATATTAAGCTTAAAGGAGCGCTGCATCA 720
Db	947 TAAATCATGCTGAAGAGAGAGGAGCGGAGTACAAATTAAGCTTAAAGGAGCGCTGCATCA 1006
Qy	721 GCTCGTTATGAACAAGAGAGTAAAAACATAGCCATGATGTTTAAATATATGTCGGTGG 780
Db	1007 GCTTGTATGAACAAGAGAGTAAAAATATAGCCATGAGCGTTTAAATATAGTCGGCGG 1066
Qy	781 AGCTGATGGCAAAACGAGGCTCATGAGGACTGACAAAAACAGAACGACGAAATGAAGATGA 840
Db	1067 AGTGGATGCAAAAGCGAGGCTCCTGGAGGACTGATAAAAACAGAACGACGAAATGAAGATGA 1126
Qy	841 GATGTACAAAATTCCTGTTGAATGATTATAGTACCGCCAGAACGAGATCCTGATCGAGAA 900
Db	1127 AATGTACAAAATTCCTGTTGAATGATTATAGTACCGCCAGAACGAGATCCTGCTCGAGAA 1186
Qy	901 CGCGAGCTGAAGAGGCTCTCCAGCAGATGAAGAAGAGAGATGATCTCTCTCTCTCTCC 960
Db	1187 TCGGAGCTGAAGAGGCTCTCCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCTCC 1246
Qy	961 TCAGAAGAGAGAGCCCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db	1247 TCAAAAG 1306
Qy	1021 TATAGAGATGACTCTGGGAACTTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db	1307 TGTAG 1366
Qy	1081 TGTAG 1140
Db	1367 TGTAG 1426
Qy	1141 AGAAAACTCGATTAACCAAGCTTCAAGAGTACACTCAGAGAGAGAGAGAGAGAGAGAG 1200
Db	1427 GGAAGAACTTGATTAACCAAGCTTCAAGAGTACACTCAGAGAGAGAGAGAGAGAGAGAG 1486
Qy	1201 CATCTCAGCAAG 1260
Db	1487 CATCTCAGCAAG 1546
Qy	1261 TAAAGAGATGATCAAGGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db	1547 TAAAGAGATGATCAAGGCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1606
Qy	1321 TGATGACACCACTCAGCTGTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db	1607 TGACGACACCACTCAGCTGTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1666
Qy	1381 AGAGAGTGGACCCCTTTTAAAG 1440
Db	1667 AGAGAGTGGTCCCTTTTAAAG 1726
Qy	1441 TACAGAGAGTGGCAATTCGATTTGGGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db	1727 TACAGAGAGTGGCTATTCGCTTTGGGTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786
Qy	1501 CTGGGTAAAGCAG 1560
Db	1787 CTGGGTAAAGCAG 1846
Qy	1561 GAAACTTTTTCAGTGCCTTTCTCAGGAAGTTCTGATCCAGACATCTTATAGTCCACTCAGC 1620

ORIGIN

Query Match	67.2%; Score 1810.2; DB 6; Length 2547;
Best Local Similarity	89.3%; Pred. No. 0;
Matches 2011; Conservative	0; Mismatches 223; Indels 17; Gaps 5;
Qy	1 CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGCAGCAGCTGAGCGGCTCTCAGGTAT 60
Db	287 CGTAGGAGAGTGACAGGAGCTGTTGTGCATGCCCGCAGCTGAACTGCTTCTCAGGGAC 346
Qy	61 CCTGGCTCTGGAACCTTGCTATGGGAGATGAGATGATGATGACAGATCAGTCTCTGTGAC 120
Db	347 CCTGGCTCTGGAGCTGGCTATGGGAGATGAGATGATGATGATGATGATGATGATGATGAT 406
Qy	121 AGAAAAACAATACTCTCAATATACCTCAGAACAGAGATGCTCCGTCAGTTTGTGA 180
Db	407 AGAAAAACAATACTCTCAATATACCTCAGAACAGAGATGCTCCGTCAGTTTGTGA 466
Qy	181 CTCGAGCAAGTCTTGCTCTTCACTACCTTTTCAAAAAAGTGCATGTTTTCGG 240
Db	467 CTCGAGCAAGTCTTGCTCTTCACTACCTTTTCAAAAAAGTGCATGTTTTCAG 526
Qy	241 TGTCTTTCGACAGAGAGAACTTGAACAAAGATTTTCTTATCTTATCAGGAGCTGAC 300
Db	527 TGCCTTCTGACAGAGAGAACTTGAACAAAGATTTTCTTATCTTATCAGGAACTGAC 586
Qy	301 CACCTTCGGGTTTCTTCTTGTATGAGATTCAGAAAGTAAAGAGCAAGAGAGAAAT 360
Db	587 TACCTTCGGTTCCTTCTTGTATGAGATTCAGAAAGTAAAGAGCAAGAGAGAGAT 646
Qy	361 AAATATAGTCTGTTTCTGAAGTATGAAAGAGTCTGCTGCTTCAAGCGAAGAACCT 420
Db	647 AAGTATAGTCTTCTTGAAGTATGAAAGAGTCTGCTGCTTCAAGCGAAGAACCT 706
Qy	421 GCTGCCAGAGAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGACATGGACCACT 480
Db	707 CTTGCCCAAGAAAGCGTGGAGACACAGAACTTGAAGCTGGGCGAGTGACATGGACCACT 766
Qy	481 GCAGAGCTGTACGCCAACTTAAGGAGCAGTTTGAAGAGCTCCAGGGCGGAGATGATCGG 540



Db 1847 GAAACTTTTCAGTGCCCTTTTCAGGAAGTTCTGTGATCCAGACAACTCTTATGATCCACCCACG 1906

Qy 1621 GCCACGGCAAGAAGCTACACAGTGGCTTAATGGGGTGCACAGCTTGCACATCAAAACT 1680

Db 1907 GCCACGGCAAGAAGCCACACAGTGTGCCTTAATGGGGTACACAGCTTGCACATCCAAACT 1966

Qy 1681 GACTAAATCTCTCTCGCTCCACCTTCTACTTTCAGACTTTCGCCAGACACATTCATGTGT 1740

Db 1967 GGCTAAGTCTCTTCGACCTCAC-----CTTCAGACTTTCGCCGCTCCTCCTCATGTGT 2020

Qy 1741 GTCTGAACACAGTTCCTCACTGCTGCTGAATATATACTCTGAAGAAAGTAAACCAAGTGA 1800

Db 2021 GTCTGAGCACAGTCCCGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2080

Qy 1801 GGTTCGAAGAAAGCAGGATCAGAAGTGAGCGTGCAGTGCAGGCCAGCTCGCGGGA 1860

Db 2081 GGTTCGAAGAAAGTACGGACCAAGAGTGAGCGTGTGTTCAGACCCAGCTCCCGGGA 2140

Qy 1861 GGGGTGCTACAGCGATGCTCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTT 1920

Db 2141 GGGTGTCTACGGTGGATGCTCTCTCGGCTCACAGCTCCACGCTGGAACGAGATGACTT 2200

Qy 1921 ACCTTAAATGTCGGGCTGCAGTGTCTGTTCCAGATGTGCGCTAGAGAGTTGACACAGG 1980

Db 2201 ACCATAGTGTCTCGGCTGCAGCGCTGT--CCTGGAGTGCAATGAGAGAAATTGACACGGG 2258

Qy 1981 GTGTAGCATAAAGTCACTGCTTAACTTAAATGCTTAAAGTCTTAAAGGATGCAAGCTTTGGTTGT 2040

Db 2259 GTGTAGCATAAAGTCACTGCTTAACTTAAATGCTTAAAGTCTTAAAGGATGCAAGCTTTGGTTGT 2100

Qy 2041 TGCTCTCCCCCAAGAGCTGAAATGCTTAAAGTCTTAAAGGATGCAAGCTTTGGTTGT 2100

Db 2318 AGTCTTTCCCCCAAGAGCTGAAAGCT-----TTAGAGATGCGAAGCTTTGGTGT 2370

Qy 2101 GTGTAGTAAACAGAGCCCTCGCTGTGACTGCAGGAATGCAATGCGTGTGGATGGAA 2160

Db 2371 GTGTAGTAAACAGAGCCCTCGCTGTGACTGCAGGAATGCAATGCGTGTGGATGGAA 2430

Qy 2161 ACAGAGCGCTGGAATGATTCCTCTG--GCCAGTACCGAAGAGACACTTTTAGGAGCTGG 2219

Db 2431 ACAAAAGCACAGAAATGATTTCTCTTCCGGGTACTGAGAAATAGCACTTTTAGGAGCTGA 2490

Qy 2220 TTCTCTAAACATTAATATTTCGTCCTCAAGT 2250

Db 2491 TTCTGTAAACATTAATTTTGTCCCAAGT 2521

RESULT 7

BC033637

LOCUS

DEFINITION BC033637 3220 bp mRNA linear PRI 27-JAN-2004

VERSION mRNA (cDNA clone MGC:45118 IMAGE:5578582), complete cds.

KEYWORDS BC033637

SOURCE MGC.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 3220)

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Schenker, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Haieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uesdin, T.B., Tothiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,

Pahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shoychenko, Y., Bouffard, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 3220)

Strausberg, R.

Direct Submission

Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps-x@mail.nih.gov](mailto:cgaps-x@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland:

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)

Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.C., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maekeri, B., Masriani, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAC Plate: 69 Row: 0 Column: 9

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7662381.

FEATURES

Location/Qualifiers

1..3220

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:45118 IMAGE:5578582"

/tissue\_type="Eye, retinoblastoma"

/clone\_lib="NIH\_MGC\_67"

/lab\_hosts="DH10B"

/notes="Vector: pCMV-SPORT6"

1..3220

/gene="SSX2IP"

/note="synonyms: KIAA0923, ADIP"

/db\_xref="GeneID:117178"

152..1996

/gene="SSX2IP"

/product="SSX2IP protein"

/protein\_id="AAH33637.1"

/db\_xref="GI:21707137"

/db\_xref="GeneID:117178"

/translation="MGDMWVTPDGLSSSEKTTISQVTSKTPSSSLYSQVLCSSIP LKSNVFFSAFCTENIQSISYLDQELTTFGPFSLYEKSKETKRELNIIVAVLNC MNELLVQRKLLAQENVEFNKLDQMDHLQSCYKLEQLETSREMIQGLQDR QLAQKRNHLQKLNKDEQVQLQNTIATSTQYNDHMKREYNKRLKHLQVME KDKKIAMDILNVYGRADGRGWSRTGKTARNEDMYKILLANDYEQKQILMENAE LKVLQOMKEMILSLSPKPKRVRVDDSTGTVISDVEBAGLSSESMDLSCTV REQLTNSIRKQWILKSHVKLQNVSKVHLEGFNEDVDVSRDHEQETKLEIQ CKEMIKTQQLQQLQATATDITTLRLRCYLLSEKERLKEENSLKEQKNERER RSTFEAIRGLRERKAFEEERASWLKQDFLNMPTTFDHQNSENVKLFSAFGSSDWNLI IVHSROPQKKPHSVGNSGVPVMSKLTLSLPASPSTSDFCQTRSCISEHSSINVLITA



ORIGIN	EBIKPNQVGEECTNQKWSVASRPGSEBQCYSCSLSYNSHVEKDLP"	Db
Query Match	63.2%; Score 1701.4; DB 5; Length 3220;	
Best Local Similarity	80.7%; Pred. No. 0;	
Matches 2214; Conservative	0; Mismatches 471; Indels 60; Gaps 17;	
QY	1 CGTAGGAGAGTACAGGAGCTGTTGTAAAGCGTCGAGCACTGAGCGCCCTCTCAGGTAT 60	QY
Db	73 CATAGAAGAGTGACAGCAGCTAGACTAAATGTTTAACTGCTGAACCTAGTTCCTCAGGTAT 132	Db
QY	61 CCTGGCTCTGGAACTTGCTATGGGAGATTGGATGACTGTGACAGATCCAGATTCGTGTAC 120	QY
Db	133 CTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTCAGATCCAGGCTCTGCTTC 192	Db
QY	121 AGAAAAAATAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCGTCACAGTTTGTGA 180	QY
Db	193 AGAAAGCAAAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA 252	Db
QY	181 CTCCAGCAAGTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGCGATGGTGTTCGG 240	QY
Db	253 CTCACAGCAAGTGTCTATGTTCTCAATACCTTTATCGAAAAATGTGCAAGTTTTTTCAG 312	Db
QY	241 TGTCTTCTGCACAGGAGAAATTTGAACAAAGTATTTCTATCTTGTATCAGAGCTGAC 300	QY
Db	313 TGCTTCTGCACAGAAGATAATTTGAACAGAGTATCTCATATCTTGTATCAGGAATTGAC 372	Db
QY	301 CACCTTCGGGTTTCTCTCTGTATCAAGAAATCCAAAAAGTAAAGAGGCAAAAGAGAGAATT 360	QY
Db	373 TACTTTTGGTTTCTCTCATATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGATT 432	Db
QY	361 AAATATAGTCGTGTTCTGAACTGTATGAAACGAGCTGTCGTGCTTCAGCGGAAAGAACT 420	QY
Db	433 AAATATAGTAGCTGTAATAATTTGTATGAATGAGCTGTTGCTTTCAGCGGAAAGAACT 492	Db
QY	421 GCTGCCCAGAGAGCGTGGAGACACAGAACTTTGAAGCTGGCAGTGACATGAGCACT 480	QY
Db	493 TCTAGCTCAGGAAATGTGGAGACAGAAATTTGAAGCTGGGAAGTATGGAACATCT 552	Db
QY	481 GCAGAGCTGTACGCAAACTTTAAGGAGCAGTTGGAACCGTCCAGCGGGAGATGATCGG 540	QY
Db	553 ACAGAGCTGTACTCAAACTTTAAGGAACAACTCGGAACCTCCAGGAGGGAATGATTGG 612	Db
QY	541 GCTTCAAGAGAGACAGCGAGCTCGAGTGCAGAAACAGAGGATTTGCAATCAGCTCCTGAA 600	QY
Db	613 GCTTCAGGAAAGAGACAGACAGTTTCAATGTAAAGAACAGGAATTTGCAATCAGCTACTAAA 672	Db
QY	601 GAAATCAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660	QY
Db	673 GAAATGAGAAAGATGGGTGCAAAAAATTTACAAAATATCATTTGCAAGTCAGCTACTCAGTA 732	Db
QY	661 TAATCATGATGTGAAGAGGAGGCGTGAATATATAAGCTTAAAGGAGCGCTGCATCA 720	QY
Db	733 TAATCATGATGAAGAGAGAAAGCGGTGAATATATAAATCTGAAGGAACGCTTACATCA 792	Db
QY	721 GCTCGTTATGAACAAGAGAGTAAAAATAGCCATGATGTTTAAATATATGGGTGCG 780	QY
Db	793 ACTTGTATGAACAAGAGAGATGAAGAAATAGCTATGGACATTTTGAATTTATGTCGGGAG 852	Db
QY	781 AGCTGATGCAAAACGAGGCTCATGAGGACTGACAAACAGAACAGCCAGGAATGAAGATGA 840	QY
Db	853 AGCTGATGGAAGAAAGAGGCTCTTGAGGAGCTGGTAAAACTGAAGCCAGGAATGAAGATGA 912	Db
QY	841 GATGTACAAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAGCAGAGATCTCTGATGGAGAA 900	QY
Db	913 AATGTATAAATTTCTTGAATGATTATGAATATCGTCAGAAACAAATCTTAATGGAAATA 972	Db
QY	901 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGAGATGATCTCTCTCTCTCTCTCC 960	QY
Db	973 TGCAGAACTTAAGAGGTTCTTCAACAAATGAAGAAAGGAATGATTTCTCTCTCTCTCTCC 1032	Db
QY	961 TCAGAGAGAGCCACGGGAAGAGCAGGACCGGCACAGGCACTGTTGCTATCTCCGA 1020	QY

1033	CCAAAGAGAAACCTAGAGAAAGAGTAGATGATAGTACAGAACTGT--TATTCCGA	1089
1021	TATAGAAGATGACTCTGGGGAACCTGAGCAGACAGACGCTGTGGGCGCTTCTCTGTGACAC	1080
1090	TGTTCAAGAGAGATGCGCGGGAACCTAAGCAGAGAGAGTAGTGTGGGACCTTCTCTGTGAAC	1149
1081	TGTGAGAGAGCAGCTGACAAACAGCATACAGGAAAACAGTGGAGAAATTTTGAAGAGTCATGT	1140
1150	TGTGAGAGAGCAGCTTACAAACAGCATCAGAAAAACAGTGGAGNAATTTTGAAAAGTCATGT	1209
1141	AGAAAACTCGATTAACCAAGCTTCCAGGTACACTCAGAGGCGCTTAAATGAGAGGACGT	1200
1210	AGAAAGCTTGATAACCAAGTTTCAAGGTACACCTCGAAGGTTTAAATGATGAAGATGT	1269
1201	CATCTCAGCAAGAACCATGAGCAAGAGACTGAGAAAACTGGAGCTGGAGATTGAGCGGTG	1260
1270	AATCTCAGCAAGAACCATGAACAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTG	1329
1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACACGCTGTGA	1320
1330	TAAAGAAATGATTAAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGTCATATGA	1389
1321	TGATGACACCACTCAGCTGTTGCGAGACTGTTACTTGTGGAAGAAAAAGGAACGCTTAA	1380
1390	TGATGATACCACTTCACTATTACGAGACTGTTATTGTTGGAGAAAAAGGAACGCTCAAA	1449
1381	AGAGAGTGGGACCTTTTAAAGACAAAAAAGAAATTTTGAGAGAGAAAAAGGCGAAGCTT	1440
1450	AGAGAAATGGTCCCTTTTAAAGACGAGAAAAAGAAATTTTGAGAGGAGAGACGAAGCTT	1509
1441	TACAGAGCTGCTCATTCGATTGGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG	1500
1510	TACAGAGCGCGTATTTCGCTGGGATTGGAGAGAAAGGCATTTGAAGAGAAAGAGCCAG	1569
1501	CTGGTAAAGCAGCAGTATTTTAAACATACAGNACTTTGACCAACCAAGAACTCAGAAATGT	1560
1570	TTGGTTAAAGCAGCAGTATTTTAAATATGACTACCTTTGACCAACCAAGAACTCAGAAATGT	1629
1561	GAACCTTTTCACTGCTGCTTCTCAGGAAGTTCCTGATCCAGACAATCTTATAGTCCATCAGC	1620
1630	GAACCTTTTCACTGCTGCTTCTCAGGAAGTTCGATTGGGACAACTTATATAGTGCATCGAG	1689
1621	GCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTGACATCAAAACT	1680
1690	GCAGCGCAAAAGAGCGCTCACAGTGTCTTAATGGGTCTCCAGTTTGCATGCTCTAAACT	1749
1681	GACTAAATCTCTTCTGCTCCTCCTTCTACCTTCTACCTTTCAGACTTTGCGCAGACATTCATGTGT	1740
1750	TACTAAATCTCTTCTGCTTCCCTTCCACTTTCAGACTTTTGCCACAGACACGTTCTCTGCAT	1809
1741	GTCTGAACACAGTTCCTCAGTGTGCTGAATATAACTCCTCGAAGAAAAGTAAACCAAGTGA	1800
1810	ATCTGACATAGTTCAATCAATGACTGTAATATAACTGCTGAGGAAATTTAAACCAATCA	1869
1801	GGTTGCAAGAGAAAGCAGGATCAGAAAGTGAAGCGTGCAGTCGAGGCCAGCTCCCGGA	1860
1870	GGTTGGAGGAGAAATGTAACAAATCAAAATGAGTGTGGCGTCAAGACCTGGATCAGAGA	1929
1861	GGGGTGTCTACAGCGATGCTCCTCGGCTTTCAGGAGCGCTCACGGGACCGAGATGACTT	1920
1930	AGGTGTCTATAGTGAATGCTCTTGGAGCTACAAAATTCATGCTAGAAAAAGAGTACTT	1989
1921	ACCTTAAATGTGCGGGTGCAGTGTGTTCCAGATGTGCCAGTACAGG-----A	1969
1990	ACCTTAGACATGTGAGCTGGAAATTTTTCATTAAATGTGTTCATCAAGTTTACATCTAA	2049
1970	GTTGACACAGGGTGTAGCATAAAGTCAGTCTCT-----TAACCTTAAGATGCTCAGAGTTG	2024
2050	GTTCAAAACAGGGTGTGTATATAAAGTCAGTTATCTCTAATAACTTAAAGATGCTCTGAGTTG	2109
2025	TTTCTTTGGACTTCTGCTCTTCCCAAGAGCTGAAATGCT--AAGCTACTTAAAGGA	2083
2110	TTTGTGGACTTCTCTCTTCCCAAGAGTTGAAATCTTAAATCTATTTTAAAGGA	2169

QY 2084 TGCAAGCTTTGG-----TTGTGTGTAGTAAACAGAGCCCTGGCTCTGTGACTGCAGGA 2139  
Db 2170 TATAAAGCTTTGGATATGTAATTTTGTAGTAACAGAGCATCTGGTCTCTGTGTAATAAGGA 2229  
QY 2140 ATGCATGCG- GTTTGGATGGAACAGAGCCCTGGAATGA--TTGCCCTCGCAGGTACCG 2196  
Db 2230 ATGTATAGATCTTTGGATGGAACAAAGAGCACTAGACTGAGTTTCCCTCTTATAGSTATTA 2289  
QY 2197 AGAAGAGACATTTAGGAGCTGGTTCCTGTGTAACATTTAAATATTTTCGTCCTCCCAAGTGTGGTT 2256  
Db 2290 AAAATAGCACATTTAGGAAATGATTAATTTGTAATAATTTTAAATTTTGTCTCAATATAGTT 2349  
QY 2257 GGCATCGAAGCTAGGCTTTACTTGAATGTATACCTGTAGATTTTAAACAAAGCAGGTTTC 2316  
Db 2350 GGCATGGAAGCTTAGCCCTTTACTTGAATGTATACCTGTAGATTTTAAACAAAGCAGGTTTC 2409  
QY 2317 TATATTTATTTAGTTAGTGTGATTTTGGGATTTACCTCTTTTCATATGTTTT- ----- 2367  
Db 2410 TATATTTATTTAGTTAGTGTG- GTTTGAAATTTACCTCTTTTCATATGTTTTTAAATAAAGT 2468  
QY 2368 -----GTGCTGTACATAAATATATACATGACTATGTTTAAGAGGCTTTTAAGGTTT 2415  
Db 2469 GAAATTTATGTATGTTTTGTACATAGATACATGATTTATGTTAAGAGGCTTTTAAGATTT 2528  
QY 2416 AAAAAGCTTACACCATGCTTCAGTATAGCATTTTCATGCCAATTTAAATGTTTTTCAGTGGC 2475  
Db 2529 AAAAGTTTCA- CAACATAGTATAGTATTTTCATGCC- AGTAAATTTTTTTTAGTGGT 2586  
QY 2476 ATGTGTTTTACAGAGG--TTAGGACCACTGCCACATGACAGTTAAGACTTTTATTTTAAAG 2533  
Db 2587 ATCTGTTTTACAGATGTTATTAGGACCAATTGATGCAATTTAAGA- ATTCTCTTTAAT 2645  
QY 2534 CCATCTGGGCAATAAATAATTCAGGCCCTTCAT- AAGCTAGTTC---AGATACTAGA 2589  
Db 2646 ACATCTGGGCAATAAATAATTTGAAAGGTATTCATGAAAGCTGAGTTCTTTAGATAATCAAC 2705  
QY 2590 ACTACTACGTTTACATTTTTCAGATTTTAAAGCAAT- GTATTTTATTTTATATATGT- G 2647  
Db 2706 ACTACTAACATTTACATTTTTCAGATTTTATGACATTTAGATTTTATTTTGTATATGTAG 2765  
QY 2648 AATGTTATTAATTTCTAAGAGGAATTTGATTTATGGAGTAATGGGG 2692  
Db 2766 AATATTATAATTTTAAAGGACTATTGATGATAGAGAAATAGGG 2810

RESULT 8  
CQ722034  
LOCUS 5835 bp DNA linear PAT 03-FEB-2004  
DEFINITION Sequence 7968 from Patent WO02068579.  
ACCESSION CQ722034  
VERSION CQ722034.1 GI:42282891  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1  
Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.  
Kites, such as nucleic acid arrays, comprising a majority of  
humanexons or transcripts, for detecting expression and other uses  
thereof  
Patent: WO 02068579-A 7968 06-SEP-2002;  
PE Corporation (NY) (US)  
Location/Qualifiers  
1..5835  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
JOURNAL  
AUTHORS  
TITLE  
FEATURES  
source  
ORIGIN  
Query Match 63.2%; Score 1701.4; DB 2; Length 5835;

Best Local Similarity 80.7%; Pred. No. 0;  
Matches 2214; Conservative 0; Mismatches 471; Indels 60; Gaps 17;  
QY 1 CGTAGGAGTGCAGAGAGCTGTTGTAAGCGTGCAGCACTGAGCCGCCCTCCTCAGGTAT 60  
Db 186 CATAGAAGAGTGCAGCAGCTAGACTAAATGTTTAACTGCTGAACTAGTTCTCCTCAGGTAT 245  
QY 61 CTTGGCTCTCGAACTTCTGCTATGGAGATTGGAGACTGTGACAGATCCAGTTCCTGTGTAC 120  
Db 246 CTTGGCTCTAGAGATTGCTATGGAGATTGGAGACTGTGACAGATCCAGTTCCTGTGTAC 305  
QY 121 AGAAAACAAAAAATCTCTCTCAATATACCTCAGAAAAAAGATGTCTCCGTCCAGTTGTGA 180  
Db 306 AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA 365  
QY 181 CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGG 240  
Db 366 CTCACAGCAAGTGTCTGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG 425  
QY 241 TGTCTTCTGCACAGGAGAGAACATTTGAACAAAATATTTCTCTATCTTGATCAGGAGCTGAC 300  
Db 426 TGCCTTCTGCACAGAGATTAATTTGAACAGATATCTCATATCTTGATCAGGAATTGAC 485  
QY 301 CACCTTCGGGTTCTCTCTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAATT 360  
Db 486 TACTTTTGGTTTTCTTCATTATATGAAGAAATCCAAAGGTAAGAGACAAAGAGAGAGTT 545  
QY 361 AATATAGTCTGCTGTTCTGAAGTATGAACGAGCTGCTGCTTCCAGCGGAAGAACCCT 420  
Db 546 AATATAGTAGTGTACTAAATTTGATGAATGAGCTGCTGTTGCTTCCAGCGGAAGAACCCT 605  
QY 421 GCTGGCCACGAGAGCGCTGGAGACACAGAACTTTGAAGCTGGGAGTGACATGGACCACCT 480  
Db 606 TCTAGCTCAGGAAAAATGTGGAGACACAGAAATTTGAAGCTGGGAGTATATGGACCATCT 665  
QY 481 GCAGAGTGTCTACGCCAAAATTTAAGAGAGAGTTTGGAAAACTGCAGCGGGAGATGATCGG 540  
Db 666 ACAGAGTGTCTACTCAAAAATTTAAGGAACAACTGGAAACCTCCAGAGGGAATGATGG 725  
QY 541 GCTTCAAGAGAGACAGAGCGAGCTGCAGTGCAGAACAGAGAGTTTGCATCAGCTCCTGAA 600  
Db 726 GCTTCAGGAAGAGACAGACAGATTTACAATGTGAAGAACAGGAATTTGCATCAGCTACTAA 785  
QY 601 GAATGAGAAAAAGAGAGTACAAAAATTAACAAATATCATAGCCAGCGGGCTACTCAGTA 660  
Db 786 GAATGAGAAAGATGAGGTGCAAAAATTTACAAAAATATCATTTGCAAGTGCAGCTACTCAGTA 845  
QY 661 TAATCATGTGTGAAGAGGAGAGCGTGAATATATAAGCTAAAGAGCGCCTGCATCA 720  
Db 846 TAATCATGATATGAAGAGAAAAAGAGCGTGAATATAATAAATGAAGGAACGCTCTACATCA 905  
QY 721 GCTCGTTTATGAACAAGAGGATAAAAACATAGCCATGATGTTTAAATTTATGTGGGTGCG 780  
Db 906 ACTTGTATTGAACAAGAAAGATTAAGAAAAATAGCAAAATAGCTATGGACATTTTGAATTTATGCGGGAG 965  
QY 781 AGCTGATGGCAAAACGAGGCTCATGAGGAGCTGACAAAAACAGAAACAGAGGAATGAAGATGA 840  
Db 966 AGCTGATGGAAAAAGAGGCTCCTGGAGGACTGGTAAAAAATGAAGCCAGGAATGAAGATGA 1025  
QY 841 GATGTACAAAATTTCTGTTGAATGATTAAGTACCGCCAGAGAGCAGATCCTGTATGGAGAA 900  
Db 1026 AATGTATAAAAATTTCTTCTTGAATGATTAATATCGTCAAGAAACAAATCTCTTAATGGAAAA 1085  
QY 901 CGCGGACCTGAAGAAGGTCTCTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCTCC 960  
Db 1086 TGCAGAACTTAAAGAGGTTCTTCAACAAATGAAGAAAGGAATGATTTCTCTCTCTCTCTCC 1145  
QY 961 TCAGAGAAGAGAGCCAGGGAAGAGAGAGGAGCGGACAGGCACTGTGTATCTCCGA 1020  
Db 1146 CCAAAAGAGAAACCTTAGAGAAAGAGTAGATGATAGTACAGGAACTGT---TATTTCGA 1202  
QY 1021 TATAGAAGATGACTCTCTGGGAACTGACAGAGACAGCGTGTGGGGCCCTTTCTCTGTGACAC 1080

Db 1203 TGTGGAAGATGCGGGGAACTAAGCAGAGAGATATGTGGACCTTTCCTGTGAAC 1262  
Qy 1081 TGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAAGTCATGT 1140  
Db 1263 TGTGAGAGAGCAGCTTACAAAACAGCATCAGAAAACAGTGGAGAAATTTTGAAAGTCATGT 1322  
Qy 1141 AGAAAACTCGATAAACCAAGCTTCGAAGGTGACACTCAGAGGGCTTTAATGAGGAGACGT 1200  
Db 1323 AGAAAACTTGATAAACCAAGCTTCAAGGGTACACCTGGGAAGGTTTAAATGATGAAGATGT 1382  
Qy 1201 CATCTCAGCAAGACCATGAGCAAGAGAGCTGAGAACTGGAGCTGGAGATTTGAGCGGTG 1260  
Db 1383 AATCTCAGCAAGACCATGAGCAAGAGAGCTGAGAACTGAGTTAGAAATTCAGCAGTG 1442  
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACACGCTGGA 1320  
Db 1443 TAAAGAAATGATTAAGAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATGCA 1502  
Qy 1321 TGATGACACCACTCAGCTGCGGAGACTGTTACTGTGCGAAGAAAGGAAAGCGCTTAA 1380  
Db 1503 TGATGATACCACTTCACTATTACGAGACTGTATTGTTGGGAAGAAAGGAAAGCGCTCAA 1562  
Qy 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1440  
Db 1563 AGAAGATGGTCCCTTTTAAAGAGCAAGAAAGAAATTTTGAGAGGAGAGACGAGCTT 1622  
Qy 1441 TACAGAAGCTGCATTCGATTCGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG 1500  
Db 1623 TACAGAAGCGCTATTTCGGCTGGGATTGGAGAGAAAGGCATTTGAAGAGAAAGAGCCAG 1682  
Qy 1501 CTGGGTTAAAGCAGAGCTTTTAAACATGACGAACTTTGACACCCAGAACTCAGAAAATGT 1560  
Db 1683 TTGGTTAAAGCAGAGCTTTCTAAATATGACTACCTTTTGACCCAGAACTCAGAAAATGT 1742  
Qy 1561 GAAACTTTTCAGTGCCTTCTCAGGAAGTCTGATCCAGACAATCTTATAGTCCACTCAGG 1620  
Db 1743 GAAACTTTTCAGTGCCTTCTCAGGAAGTCTGATTTGGGACAATCTTATAGTCCACTCAG 1802  
Qy 1621 GCCAGGCAAAAGAGCTACACAGTGTGGCTTAATGGGTTGCAGCTTGACATCAAAACT 1680  
Db 1803 GCAGCGCAAAAGAGCGCTCACAGTGTGTCTAATGGGCTCCAGTTTGCATGCTCAAACT 1862  
Qy 1681 GACTAAATCTCTCTGCTCCTACCTTCTACTTCAGACTTTCGCCAGACACATTCATGTGT 1740  
Db 1863 TACTAAATCTCTCTGCTTCTACCTTCCACTTCAGACTTTTGCCAGACACGTTCTCGCAT 1922  
Qy 1741 GTCTGAACACAGCTTCCATCAGTGTGCTGAATATACTCTGAAGAAAGTAAACCAAGTGA 1800  
Db 1923 ATCTGAACATAGTTCATCAATGATGACTGATATATACTGCTGAGAAATTTAAACCAATCA 1982  
Qy 1801 GGTTCGAAGAGAAACGAGATCAGAAAGTGGAGCGTGCAGTGCAGGCCAGCTCGCGGA 1860  
Db 1983 GGTTCGAGGAGAACGTACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGGA 2042  
Qy 1861 GGGGTGCTACAGCGATGCTCTCGGCTTCAGGAGCGCTACGGGGACCGAGATGACTT 1920  
Db 2043 AGGTTGCTATAGTGGATGCTCTCTGAGCTACAAAATTTCTCATGTAGAAAAAGATGACTT 2102  
Qy 1921 ACCTTAAATGTGCGGGCTGCAGTGTGTTCCAGATGTGCGCTAGAGG-----A 1969  
Db 2103 ACCTTAGACATGTGAGCTGGAAATTTTTCATTAATGTGTTTCATCAAGTTTCACATCTAA 2162  
Qy 1970 GTTCAGCAGGGTGTAGCATAAAGTCAGTCGTC-----TAACTTAAGATGCTCAGAGTTG 2024  
Db 2163 GTTGAACAGGGTGTGTATAAAGTCAGTTATCTCTTAATACTTAAGATGGTCTGAGTTG 2222  
Qy 2025 TTTGTTTCGAGCTTCGCTGCTCTCCCAAGAGCTGAAATGCT-NAGCTACTTAAAGGA 2083  
Db 2223 TTTGTTTGAGCTTCCTGCTCTCTCCCAAGAGGTTGAAATCTTAAATCTATTTAAAGGA 2282  
Qy 2084 TCGAAAGCTTTGG-----TTGTGTGTGTAGTAAACAGAGCCCTGGCTCTGTGACTGCAGGA 2139  
Db 2283 TATAAAGCTTTGGATATGTATTTTAGTAAACAGAGCATCTGGTTCTGTGAATAAGGA 2342

Qy 2140 ATGCATGGC-GTTTGGATGGAAACAGAAAGCGCTGGAATGA--TTGCCTCGCCAGGTACCG 2196  
Db 2343 ATGTATAGATGTTTGGATGGAAACAAAGCACTAGACTGAGTTTCTCTTATAGGTATTA 2402  
Qy 2197 AGAAGAGCACTTTTAGGAGCTGCTTCGTGTAACATTAATAATTCGTCCCAAGTGTGGTT 2256  
Db 2403 AAAATAGCACTTTTAGGAACTGATTAATTGTAATGTTTAAATTTTGTCTCAAAATAGTT 2462  
Qy 2257 GCATTGGAAGTGTAGCCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTTC 2316  
Db 2463 GCATTGGAAGTGTAGCCCTTTACTTGAATGTATCTGTAGATTTTAAACAAAGCAGGTTTC 2522  
Qy 2317 TATATTTATTAATGTTTAGTGTGATTTTGGGATTTACCTCTCTTTCATATGTTTT----- 2367  
Db 2523 TATATTTATTAATGTTTAGTGTG-GTTTGAATTTACCTCTTTCATATGTTTAAATAAAGT 2581  
Qy 2368 -----GTGCTGTACATAAATATATACATGACTATGTAAAGAGCTTTTAAGGTTT 2415  
Db 2582 GAAATTTATGATGTTTGTACATAGATACATGATTAATGTTAAAGAGCTTTTAAGGTTT 2641  
Qy 2416 AAAAACTTTCACACCATGCTTGAGTATAGCATTTTCATGCCAATTAATAATGTTTTCAGTGGC 2475  
Db 2642 AAAAGTTTCACA-CAACCATAGTATAGTATTTTCATGCC-AGTAAATTTTTCAGTGGT 2699  
Qy 2476 ATGTGTTTTCACAGAGG--TTAGGACCACTGCCACATGACAGTTAAAGAGCTTTTAAAG 2533  
Db 2700 ATTCTGTTTACAGATGTTATAGGACCATGATGCAATTAATTAAGA-ATTCTCTTAAT 2758  
Qy 2534 CCATCTGGGCAATAAAATTCAAAGCCCTTCAT-AAGCTGAGTTC---AGATAACTAGA 2589  
Db 2759 ACATCTGGGCAATAAATTTGAAAGGTATTCATGAAGCTGAGTTCCTTTAGATAATCAAC 2818  
Qy 2590 ACTACTACGTTTACATTTTTCAGATTTTAAAGCAAT--GTATTTTATTTATATATGT-G 2647  
Db 2819 ACTACTACATTCACATTTTTCAGATTTTATGACATTAAGATTTTATTTTGTATATGTAG 2878  
Qy 2648 AATGTTATTAATTTCTAAGAGGAATATTGATTATGGAGTAATGGGG 2692  
Db 2879 AATATTATAATTTTAAAGGACTATTGATGATAGAAGAAATAGGG 2923  
RESULT 9  
AB023140  
LOCUS Homo sapiens mRNA for KIAA0923 protein, partial cds. PRI 10-JAN-2004  
DEFINITION  
ACCESSION AB023140  
VERSION AB023140.1 GI:4589477  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE  
1 Nagase, T., Ishikawa, K., Suyama, M., Kikuno, R., Hirotsawa, M.,  
Miyajima, N., Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.  
Prediction of the coding sequences of unidentified human genes.  
XIII. The complete sequences of 100 new cDNA clones from brain  
which code for large proteins in vitro  
DNA Res. 6 (1), 63-70 (1999)  
JOURNAL PUBMED 10231032  
REFERENCE 2 (bases 1 to 5835)  
AUTHORS Ohara, O., Nagase, T. and Kikuno, R.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1999) Osamu Ohara, Kazusa DNA Research Institute,  
Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba  
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,  
Fax:+81-438-52-3914)  
FEATURES  
Location/Qualifiers  
1..5835  
/organism="Homo sapiens"  
/mol\_type="mRNA"

gene	/db_xref="taxon:9606"	
	/clone="hh02882"	
	/sex="male"	
	/tissue_type="brain"	
	/clone_lib="pbluescriptII SK plus"	
	/dev_stage="adult"	
	1..5835	
	/gene="KIAA0923"	
	<229..2109	
	CDS	/gene="KIAA0923"
/note="Start codon is not identified"		
/codon_start=1		
/product="KIAA0923 protein"		
/protein_id="BAA76767.2"		
/db_xref="GI:40788987"		
/translation="TSSSGILALEIAMGDMVTVDPGLSSSEKTSIQVTSKTMSPSS LYSQVCSSTPLSKNVHSPFAFCTEDNIEQISYLDQELTTFGFSLSYBESKGET KRLNIVAVLNCMVELLVQRNLLAQENVETONKLSQDMHQLQSCYSLKLEQLETS RREMIGLQERDQLQCKNRNLHLLKNEKDEVOKLONI IASRATQYNDHMKREREYN KLKRLHLQVNVKDKKIAMDLINYVGRADGKRGSMRTGKTEARNEDEMYKILLNDYE YRQQLMENALELKVLOOMKKEMISLLSPQKKPRPRVDDSTGTVISDVEEDAGELS RESMDUGSCETVRQLTNSIRKQWRIILKSHVEKLDNQVSKVHLEGFNDEEDVISQDHE QETEKLELEIQCKEMIKTQOQLLQQQLATAYDDDTLSLLRDCYLLBEKERLKEENSL PFEQKNFERERRSFTAAIRLGLKAFEEERASWLKQOFLNNMTTFDHNSENKSLF SAFGSSDWDNLI VHSRQPKPHSVNSGSPVCMKLTQSLPASPSPTDFOCTRSCIS EHSNIVNLITABEIKPNQVGECTNQKWSVASRPSGQEGCYSCSLSLYTNSHVEKDD LP"		
ORIGIN		
Query Match 63.1%; Score 1699.8; DB 5; Length 5835;		
Best Local Similarity 80.6%; Pred. No. 0;		
Matches 2213; Conservative 0; Mismatches 472; Indels 60; Gaps 17;		
1 CGTAGAGAGTGACAGGAGCTGTTGTAAGCGTCGCAGCAGCTGAGCGGCTCTCTCAGGTAT 60		
186 CATAGAAGAGTGACAGCAGCTAGACTAAATGTTTAACTGCTGCACTAGTTCTCTCAGGTAT 245		
61 CCTGGCTCTGGAACCTTGCTATGGGAGATTGATGACTGTGACAGATCCAGTTCTGTGTAC 120		
246 CCTGGCTCTAGAGATTGCTATGGGAGATTGATGACTGTGACAGATCCAGGTCTGTCTTC 305		
121 AGAAAAACAAAATCTCTCTCAATATACCTCAGAAAACAAAGATGTCTCCGTCACGTTTGTGA 180		
306 AGAAGCAAACTATCTCTCAATATACCTCAGAAAACAAAGATGTCTCCATCAAGTTTATA 365		
181 CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAACGTCGATGGTGTGTTTCGG 240		
366 CTCACAGCAGTGCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAG 425		
241 TGTCCTCTGCACAGAGAGAACATTGAAACAAAGTATTTCTCTATCTTCATCAGGAGCTGAC 300		
426 TGCCTTCGACAGAGATATATTTGAACAGAGTATCTCATATCTTGATCAGGAATTGAC 485		
301 CACCTTCGGGTTTCTCTCTTGATGAAGAAATCCAAAAGTAAAGAGGCAAAAGAGAGAATT 360		
486 TACTTTTGGTTTCTCTTCATTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGATT 545		
361 AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTGTGCTTCAGCGGGAAGAACT 420		
546 AAATATAGTAGCTGTACTAAATTTGATGAATGAGCTGCTGTTGTGCTTCAGCGGGAAGAACT 605		
421 GCTGGCCACAGAGAGCGTGGAGACACAGAACTTGAAGCTGGCAGTGACATGGACCACT 480		
606 TCTAGCTCAGGAAATGTGGGAGACAGAAATTTGAAGCTGGGAAGTATATGGACCACT 665		
481 GCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTTGGAAACGTCCAGCGGGAGATGATCGG 540		
666 ACAGAGCTGCTACTCAAACTTAAGGAACAACCTGGAAACCTCCAGGAGGGAATGATTGG 725		
541 GCTTCAAGAGAGACAGGAGCTGCAGTGAAGNACAGGAGTTTGCATCAGCTCCTGAA 600		
726 GCTTCAGAAAGAGACAGACAGTTTACAAATGTAAGNACAGGAATTTGCATCAGCTACTAAA 785		

Qy	601	GAATGAGAAAGATGAGGTACAAAAATTTACAAAAATATCATAGCCAGCCGGCTACTCAGTA 660
Db	786	GAATGAGAAAGATGAGGTGCAAAAATTTACAAAAATATCATTTGCAAGTGAAGTACTCAGTA 845
Qy	661	TAATCATGATGTGAAGAGGAGGCGTGAATATAATAAGCTAAAGAGGCGCCTGCAATCA 720
Db	846	TAATCATGATATGAAGAGAAAGAGCGTGAATATAATAAACTGAAGGAAACGCTACATCA 905
Qy	721	GCTCGTTATGAACNAGAGGATAAAAACATAGCCATGGATGTTTTAAATATATGCGGTCG 780
Db	906	ACTTGTATGAACNAGAAAGATAGAAAAATAGCTATCGACATTTTGAATATATGTCGGGAG 965
Qy	781	AGCTGATGGCAACGAGGCTCATCGAGGACTGACAAAAACAGAAAGCCAGGAATGAAGATGA 840
Db	966	AGCTGATGGAAAAAGAGGCTCTCGAGGACTGGTAAAACTGAAGCCAGGAATGAAGATGA 1025
Qy	841	GATGTACAAAAATCTGTTGAATGATTAATAGTACCCGCAAGAGCAGATCTCTCTCTCTCC 900
Db	1026	AATGTATAAAAAATCTCTTGAATGATTAATGAATATCGTCAGAAAACAAAATCCTAATGGA 1085
Qy	901	CGCGGAGCTGAAGAGTCTCTCAGCAGATGAGNAGAGAGATGATCTCTCTCTCTCTCTCC 960
Db	1086	TGCAGAACTTAAAGAGGTTCTTCAACAAATGAAAAAGAAATGATTTCTCTCTCTCTCTCC 1145
Qy	961	TCAGAAAGAAAGCCAGGGAAGAGCAGAGGACGGCACAGGCACTGTGCTATCTCCGA 1020
Db	1146	CCAAAGAAAGAACCTTAGAAGAGATGATGATAGTACAGGAACCTGT---TATTTCCGA 1202
Qy	1021	TATAGAAAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCTCTTCTCTGTGACAC 1080
Db	1203	TGTTGAAGAAAGATCGCCGGGAACTTAAGCAGAGAGATATGTGGGACCTTTCTCTGTGAAC 1262
Qy	1081	TGTGAGAGACAGCTGACAAACAGCATCAGNAAACAGTGGAGATTTTGAAGAGTCAATGT 1140
Db	1263	TGTGAGAGACAGCTTACAAAACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCAATGT 1322
Qy	1141	AGAAAACTCGATAAACCAAGCTTCGAAGGTACACTCAGAGGGGCTTAAATGAGGAGGACGT 1200
Db	1323	AGAAAGCTTGATAAACCAAGTTTCAAGGTACACCTCGNAGGTTTAAATGATGAAGATGT 1382
Qy	1201	CATCTCAGCAAGACCATGAGCAAGAGACATGAGAAACCTGGAGCTGGAGATTGAGCGGTG 1260
Db	1383	AATCTCAGCAAGACCATGAACAAAGAACTGAAAACTCGAGTTAGAAAAATTCAGCAGTG 1442
Qy	1261	TAAAGAGATCATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGCGCCACACCGTGTGA 1320
Db	1443	TAAAGAAATGATTAAGAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGATATGA 1502
Qy	1321	TGATGACACCACTCACTGTTCGAGACTGTTACTTCTCGAAGAAAGAAAGGAAACGCTTAA 1380
Db	1503	TGATGATACCACTTCACTATTACGAGACTGTTATTTGTTGGAAGAAAGGAAACGCTCTCAA 1562
Qy	1381	AGAAGAGTGAACCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGGCGAAAGCTT 1440
Db	1563	AGAAGAAATGCTGCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGGAGAGAGCAAGGCTT 1622
Qy	1441	TACAGAGAGCTGCCATTTCGATTGGGGTTGGAGAGAAAGCGCTTTGAAGAAAGAGCGGACG 1500
Db	1623	TACAGAAAGCGCTATTTCGCTGGGATTGGAGAGAAAGGCAATTTGAAGAGAAAGAGCGGACG 1682
Qy	1501	CTGGGTAAGACAGCAGTTTTTAAACATGACGAACTTTTGACCACAGAACTCAGAAAAATGT 1560
Db	1683	TTGGTTAAAGCAGCAGTTTTCTAAATATGACTACCTTTGACCACCCAGAACTCAGAAAAATGT 1742
Qy	1561	GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAGTCCACTCAGC 1620
Db	1743	GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATGGGACAAATCTTATAGTGCACCTGAG 1802
Qy	1621	GCCACGCGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTTGACATCAAAACT 1680
Db	1803	GCAGCGCGCAAAAGAGCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTAACT 1862
Qy	1681	GACTAAATCTCTTCTGCTCCATCCTTCTACTTTCAGACTTTTCGCCAGACACATTCATGTGT 1740

Db 1863 TACTAAATCTCTCTGCTTACCTTCCACTTCAGACTTTTCCAGACACAGTTCCTGCAAT 1922  
Qy 1741 GTCTGAACACAGTTCATCAGTGTGCTGAATATACTCTGAAGAAAGTAAACCAAGTGA 1800  
Db 1923 ATCTGAACATAGTTCAATCAATGTACTGAATATACTGTGAAGAAATTAACCAATCA 1982  
Qy 1801 GGTTCGAAGAAACACGATCAGAGTGGAGCGTGCAGTCGAGCGCCAGTCTCGGGA 1860  
Db 1983 GGTTCGAGAGAAATGTACAAATCAAAATGGAGTGGGCATCAAGACCTGGATCAGGA 2042  
Qy 1861 GGGGTGTACACGGATGCTCTCGGCCCTTCAGGAGCGCTCACGGGACCGAGATGACTT 1920  
Db 2043 AGTTTGTCTATAGTGTCTCTTGAGCTACAAATCTCATGTAGAAAGATGACTT 2102  
Qy 1921 ACCTTAAATGTGGGCTGCACTGCTGTTCAGATGTCGCGTAGAGG-----A 1969  
Db 2103 ACCTTAGACATGTGACTGGAATTTTTTTCATTAATGTTCATCAAGTTTCACATCTAA 2162  
Qy 1970 GTTGACACAGGGGTAGCATAAAGTCAGTCGTC-----TAACCTTAAGATGCTCAGATG 2024  
Db 2163 GTTGAACAGGGTGTGTATAAAGTCAGTTATCTCTAATACTTAAGATGGTCTGAGTTG 2222  
Qy 2025 TTTGTTTGACATCGCTGCTCTCCCAAGAGCTGAAATGCT-AAGCTACTTAAAGGA 2083  
Db 2223 TTTGTTTGACATCGCTGCTCTCCCAAGAGTGGAAATCTTAATCTTAATTAAGGA 2282  
Qy 2084 TGCAAAGCTTTGG-----TTGTGTGTAGTAACAGAGCCCTGGCTCTGTGACTGCAGGA 2139  
Db 2283 TATAAAGCTTTGGATATGTATTTTGTAGTAACAGAGCATCTGTTCTGTGAATAAGGA 2342  
Qy 2140 ATGCATGCG-GTTTGGATGGAACAGAGCGGTGGAATGA--TTGCCTCGCCAGGTACCG 2196  
Db 2343 ATGTATAGATGTTGGATGGAACAAAGCACTAGACTGAGTTTCCTCTTATAGGTATTA 2402  
Qy 2197 AGAAGACACTTTTAGGACTGGTTCCTGTAACATAATAATTCGTCGCCAGTGTGTT 2256  
Db 2403 AAATAGACACTTTTAGGAACCTGATTTGTAAATGTTTAAATTTTGTCTCAATATAGTT 2462  
Qy 2257 GGCAATGGAAGTGTAGCCCTTTACTTGAATGTATCTGTAGATTTTAAACAAGCAGGTTTC 2316  
Db 2463 GGCATTGGAAGTTTAAAGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAGCGAGTTTC 2522  
Qy 2317 TATATTTATATGTTTAGTGTGATTTTGGGATTTACCTCTTTCATATGTTTT----- 2367  
Db 2523 TATATTTATATGTTTAGTGTG-GTTTGAATTTACCTCTTTCATATGTTTTTAAATAAGT 2581  
Qy 2368 -----GTCTGTACATAAATATACATGACTATGTAAAGAGCTTTTAAAGTTT 2415  
Db 2582 GAAATTTATGATGTTTTTGTACATAGATACATGATTTATGTTAAGAGCTTTTAAAGTTT 2641  
Qy 2416 AAAAATTTTACACCATGCTTGAAGTATAGCATTTTCATGCCAAATTTAAATGTTTTTTCAGTGGC 2475  
Db 2642 AAAAGTTTTCACA-CAACCATAGTATGATTTTTCATGCC-AGTAAATTTTTTTAGTGT 2699  
Qy 2476 ATGTTGTTTACAGG--TTAGGACCACTGCCCATGACAGTTAAGACTTTATTTTAAAG 2533  
Db 2700 ATCTGTTTACAGATGTTATAGGACCAATGATGCAATTACATTTAAGA-ATTCTCTTTAAT 2758  
Qy 2534 CCATCTGGGCAATAAATAATCAAGCCCTTCAT-AAGCTGAGTTC---ACATAACTAGA 2589  
Db 2759 ACACTGGGCAATAAATAATGGAAGGTATTCATGAAGCTGAGTTCTTTAGATATATCAAC 2818  
Qy 2590 ACTACTAAGCTTTACATTTTGGAGATTTTAAAGCAATT-GTATTTTATTTTATATATGT-G 2647  
Db 2819 ACTACTAATACATTTTGGAGATTTTATGACATATAGATTTTATTTTGTATATGTAG 2878  
Qy 2648 AATGTTTAAATTTCTAAGAGGAATATGATTTATGGAGTAATGGGG 2692  
Db 2879 AATATTATAATTTTAAAGGACTATTGATGATGAAGAATAGGG 2923

AY651262  
LOCUS Homo sapiens synovial sarcoma X breakpoint 2 interacting protein  
DEFINITION (SSX2IP) mRNA, partial sequence.  
ACCESSION AY651262  
VERSION AY651262.1 GI:60686984  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Guinn,B.A., Bland,E.A., Lodi,U., Liggins,A.P., Tobal,K.,  
Peters,S., Wells,J.W., Banham,A.H. and Mufti,G.J.  
TITLE Humoral detection of leukaemia-associated antigens in presentation  
acute myeloid leukaemia  
JOURNAL Biochem. Biophys. Res. Commun. 335 (4), 1293-1304 (2005)  
PUBMED 16112646  
REFERENCE  
AUTHORS Guinn,B.A., Bland,E.A. and Mufti,G.J.  
TITLE Direct Submission  
JOURNAL Submitted (13-JUN-2004) Department of Haematological Medicine,  
Guy's, King's & St. Thomas' School of Medicine, 123 Coldharbour  
Lane, London SE5 9NU, U.K.  
FEATURES  
Location/Qualifiers  
1..3073  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p22.3"  
<1..3073  
/gene="SSX2IP"  
misc\_feature <1..1649  
/gene="SSX2IP"  
/note="similar to synovial sarcoma X breakpoint 2  
interacting protein; contains possible frameshift"  
ORIGIN  
Query Match 56.2%; Score 1512.2; DB 5; Length 3073;  
Best Local Similarity 80.4%; Pred. No. 0;  
Matches 2023; Conservative 0; Mismatches 428; Indels 64; Gaps 19;  
Qy 231 GTGTTTTCGGTGTCTTCTGCACAGGAGAACATTGAACAAAGTATTTTCTTATCTTGATC 290  
Db 1 GTTTTTTCAGTGCCTTCTGCACAGAGATAATATTGAACAGAGTATCTCATATCTTGATC 60  
Qy 291 AGGAGCTGACCACCTTCGGGTTTCTCTTCTTGTATGAAGATCCAAAAGTAAAGAGCAA 350  
Db 61 AGGAATTTGACTACTTTTGGTTTTTCTTCTTATATGAAGATCCAAAAGGTAAGAGACA 120  
Qy 351 AGAGAGAATTAATATAGTCGCTGCTTCTGAACCTGTATGAAGAGCTGCTGCTTTCAGC 410  
Db 121 AGAGAGATTAATATATAGTACTTAATTTGTATGAATGAGCTGCTGCTGCTTTCAGC 180  
Qy 411 GGAAGAACCTCTGCTGGCCCGAGAGCGTGGAGACACAGAACTTGAAGCTGGGCACTGACA 470  
Db 181 GGAAGAACCTTCTAGCTCAGGAAATGTTGGAGACACAGAAATTTGAAGCTGGGAAGTATA 240  
Qy 471 TGGACCACTGCAGAGCTGCTACGCCAACTTAAGGAGAGTGGAAACGTCCAGGCGGG 530  
Db 241 TGGACCACTGCAGAGCTGCTACTCAAAACTTAAAGAACCAACTGGAACCTCCAGGAGGG 300  
Qy 531 AGATGATCGGGCTTCAAGAGAGACAGAGCGAGCTGCAGTGCAGAACACAGAGATTGTCATC 590  
Db 301 AAATGATTGGCTTTCAGAAAGAGACAGACAGTGTACAAATGTAAAGAACAGGAATTTGCATC 360  
Qy 591 AGCTCTCTGAAGAATGAAGAAGATGAGGTACAAAAATATCAAAAAATATCATAGCCAGCCGG 650  
Db 361 AGCTACTTAAGAATGAAGAAGATGAGGTGCAAAAATATCAAAAATATCATTTGCAAGTCGAG 420  
Qy 651 CTACTCAGTATATCATGTGTGAAGAGGAGCGTGAATATAATTAAGCTAAAGAGGC 710

|||||  
421 CTACTCAGTATATCATGATATGAAGAGAAAGAGCGTGATATATAATANAACCTGAGGAGAC 480  
QY  
711 GCCTGCACTACGCTCGTTATGAAACAAAGAGGATAAAAAACATAGCCATGGATGTTTAAATT 770  
Db  
481 GTCTACATCAACTGTTTATGAACAAGAAAGATAAGAAAAATAGCTATGGACATTTTGAATT 540  
QY  
771 ATGTGGGTGAGCTGTAGTGGCAAAACGAGGCTCATGGAGGACTGACAAAACAGAACGACCGAGGA 830  
Db  
541 ATGTGGGAGAGCTGATGGAAGAAAGAGGCTCCTGGAGGACTGGTAAACCTGAAGCCAGGA 600  
QY  
831 ATGAAGATGAGATGTACAAAATCTCTTTGAATGATTTATGATACCGCCAGAGCAGATCC 890  
Db  
601 ATGAAGATGAATGTATAAAATCTCTTGAATGATTTATGAATATCGTCAGAAAACAAATCC 660  
QY  
891 TGATGGAGAACCGGAGCTGAAGAGGTCTCCAGCAGATGAAGAAGAGATGATCTCTC 950  
Db  
661 TAAATGGAAAATGCAGAACTTAAGAAGGTTCTTCAACAAATGAAAAGAGMAATGATTTCTC 720  
QY  
951 TCCTGTCTCCTCAGAAAGAGAGCCAGGGAAGAGAGAGGACGGCAGACGCACTCTTG 1010  
Db  
721 TTCCTTCTCCCAAAAGAGAAACCTAGAGAAAGAGTAGATGATACAGGAACACTGT-- 778  
QY  
1011 CTATCTCCGATATAGAAGATGACTCTGGGGAACCTGAGCAGACAGAGCGTGTGGGCGCTTT 1070  
Db  
779 -TATTTCCGATGTTGAAGAGATCGCGGGAACCTAAGCAGAGAGATGATGTGGACCTTT 837  
QY  
1071 CCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTGA 1130  
Db  
838 CCTGTGAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAAACAGTGGAGAAATTTGA 897  
QY  
1131 AAAGTCATGTAGAAAACCTCATACCAAGCTTCGAAAGGTACACTCAGAGGCGCTTAATG 1190  
Db  
898 AAAGTCATGTAGAAAAGCTTGATACCAAGTTTCAAAGGTACACTGGAAGGTTTTAATG 957  
QY  
1191 AGGAGAGCGTCATCTCAGCAAGACCATGAGCAAGAGACTGAGAAAACCTGAGCGTGAGA 1250  
Db  
958 ATGAAGATGTAATCTCACGACAGACCATGAACAGAACTGAAAACCTCGATTAGAAA 1017  
QY  
1251 TTGAGCGGTGAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCA 1310  
Db  
1018 TTCAGCAGTGAAGAAATGATTAATACTCAGCAACAGCTTTTA--CAGCAGCTCGCTA 1074  
QY  
1311 CCAGTGTGATGATGACACCACTCAGCTGTTGCGAGACTGTACTGCTGGAAGAAAGG 1370  
Db  
1075 CTGCATATGATGATACCACTTCCACTATTTACGAGACTGTTATTTGTTGGAGAAAGG 1134  
QY  
1371 AACGCTTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAA 1430  
Db  
1135 AACGCTCTCAAGAGAAATGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGAGGAGA 1194  
QY  
1431 GCGGAAGCTTTACAGAACTCGCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAAG 1490  
Db  
1195 GACGAAGCTTTACAGAAAGCGCTATTCGCTGGGATTTGGAGAGAAAGCAATTTGAAGAAG 1254  
QY  
1491 AGCGAGCCAGCTGGTAAAGCAGCAGTTTAAACATGACCACTTTTGACCAACAGCAACT 1550  
Db  
1255 AAAGAGCCAGTTGGTTAAAGCAGCAGTTTCTAAATATGACTACCTTTTGACCAACAGCAACT 1314  
QY  
1551 CAGAAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAAATCTTATAG 1610  
Db  
1315 CAGAAAATGTGAAAATCTTTTCAGTGCCTTCTCAGGAAGTTCTGATTTGGGACAAATCTTATAG 1374  
QY  
1611 TCCACTACGGCCAGCGGAAAAGAGAGCTACACAGTGTGGCTTAATGGGTCAGCTTGCA 1670  
Db  
1375 TGCACCTCAGGCGAGCGGAAAAGAGAGCTCACAGTGTGTCTAATGGGTCCTCCAGTTTGA 1434  
QY  
1671 CATCAAAATGCAATAAATCTCTCTGCTCACCTTCTACTTCTCAGACTTTCCGACAGACAC 1730  
Db  
1435 TGTCTAAACTTACTTAAATCTCTCTCTGCTTCCCTTCCACTTCCACTTTCAGACTTTTGGCAGACAC 1494  
QY  
1731 ATTCAATGTGTCTGAACACAGCTTCCATCAGTGTGCTGAATATAACTCTCTGAAGAAAGTA 1790  
|||||

Db  
1495 GTTCTCGCATATCTGAACATAGTTCAATCAATGCTGTAATAATAAAGCTGCTGAAGAAATTA 1554  
QY  
1791 AACCAAGTGAAGTTGCAAGAGAAAGCACGGATCAGAGTGGAGCGTGCAGTCGAGGCCCCA 1850  
Db  
1555 AACCAAAATCAGGTTGGAGGAGAACGTACAAATCAAAATCAAAATCGAGTGTGGCGCTCAAGACCTG 1614  
QY  
1851 GCTCGGGGAGGGGTGCTACAGCGGATGCTCTCGGCCCTTCAGAGAGCGCTCAGCGGAGCC 1910  
Db  
1615 GATCAGAGGA-AGTGCTATAGTGGATGCTCTTGGAGTACACAAATCTCATGTAGAAA 1673  
QY  
1911 GAGATGACTTACCTTAAATGTGCGGGCTGCAGTGTCTTCCAGATGTGCGGTAGAGG-- 1968  
Db  
1674 AAGATGACTTACCTTGAACATGTGGACTGGAAATTTTTTTCATTAATGTGTTCATCAAGTT 1733  
QY  
1969 -----AGTTGACACAGGCTGTAGCATAAAAGTCAAGTCAAGTCTGCTC-----TAACTTAAAGATG 2014  
Db  
1734 TCACATCTAAGTTTGAACACAGGGGTGTGTATATAAAGTCAAGTTATCTCTAATAACTTAAAGATG 1793  
QY  
2015 CTGAGAGTCTTGTGTTTGGACTTTCGCTGTCTTCCCCCAAGAGCTGAAATGCT--AAGCTA 2073  
Db  
1794 GTCTGAGTGTGTTTGGACTTTCCTGTCTTCCCCCAAGAGTGTGAAATCTTAAATCTA 1853  
QY  
2074 CTTAAAGAGATGCAAAAGCTTTGG-----TTGTGTGTAGTAAACAGAGCCCTGCTCTGT 2129  
Db  
1854 TTTAAAGAGATATAAAAGCTTTGGATATGATATTTTGTAGTAAACAGAGCATCTGGTCTGT 1913  
QY  
2130 GACTGCAGGAATGCAT--GGCGTTTGGATGGAACAGAGCGCTGGAGTGA--TTGCGCTCG 2186  
Db  
1914 GAATAAGGAATGTATAGATGTTTGGATGGAACAAAGCAGCTAGACTGAGTTTCTCTTT 1973  
QY  
2187 CAGGTAACCCAGAGAGACACTTTTAGGGACTGCTTCTGTAAACATTTAAATATTTGCTCCC 2246  
Db  
1974 ATAGTATTTAAATAAGTACTTTTAGGAACTGATTTTGTAAATGTTTAAATTTGTCTC 2033  
QY  
2247 AAGTGTGTTGGCAATTTGGAAGTGTAGCTTTTACTGAAATGTATATCTGATAGATTTTAAACA 2306  
Db  
2034 AAATATAGTTGGCAATTTGGAAGTTTAGCTTTTACTTGAATGTATATCTGATAGATTTTAAACA 2093  
QY  
2307 AAGCAGGTTCTATATTTATTTATGTTAGTGTGATTTTGGGATTTTACCTTTTTCATATCTTT 2366  
Db  
2094 AAGCGAGTTCTATATTTATTTATGTTTAGTGTG--GTTTGAATTTTACCTTTTTCATATGTTT 2152  
QY  
2367 T-----GTGCTGTGTACATAAATATACATGACTATGTTTAAGAGGC 2405  
Db  
2153 TAAATAAGTGAAATTTATGATGTTTGTGTACATAGATACACATGATTTATGTTAAGAGGC 2212  
QY  
2406 TTTAAGGTTTAAAACTTTACACACATGCTTGTAGTATAGCAATTTATGCTCCCAATTTAAATGT 2465  
Db  
2213 TTTAAGATTTTAAAGTTTTCACA--CAACCATAAAGTATAGTATTTTCATGCC--AGTAAAAATTT 2270  
QY  
2466 TTTTCAGTGGCATGGTGTTTTACAGAG--TTTAGGACCACTGCCACATGACAGTTAAGACTT 2523  
Db  
2271 TTTTAGTGGTATTTCTGTTTACAGATGTTATTAGGACCAATTTGATGATTTACATTTAAGA--AT 2329  
QY  
2524 TATTTTAAAGCCATCTGGGCAATAAAATTTCAAGGCCCTTTCAT--AAGCTGAGTTC--A 2579  
Db  
2330 TCTCTTTAATACATCTGGGCAATAAATTTTGAAGGTATTTCCATGAAGCTGAGTTCTTTA 2389  
QY  
2580 GATAACTAGAACTACTAACGTTTACATTTTGTAGATTTTAAAGCAATTTGATTTTATTTT 2638  
Db  
2390 GATAATCAACACTACTAACATTTTGTAGATTTTGTAGATTTTGTAGATTTTGTATTTT 2449  
QY  
2639 ATATATGT--GAATGTTATATAATTTCTAAGAGGAATATTGATTTAGAGTAAATGGG 2692  
Db  
2450 GTATATGTAGAAATATATAAATTTTAAAGGACTATTGTATGATAGAGAATAGGG 2504  
|||||

RESULT 11

AY367055

LOCUS

DEFINITION

ACCESSION

AY367055 2379 bp mRNA linear PRI 09-SEP-2003  
Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein  
(SSX2IP) mRNA, complete cds.

AY367055.1	GI:34451652																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			</
------------	-------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----



```
Qy 1441 TACAGAGCTCCATTGATGGGTTGGAGAGAAAGCGGTTTGAAGAGCGAGCCAG 1500
Db 1548 TACAGAGCGCCCTATTCCGCTGGGATTGGAGAGAAAGCGCATTTGAAGAGAAAGAGCCAG 1607
Qy 1501 CTGGGTAAGCAGCAGATTTTAAACATGACCAACTTTGACCACCAAGAACTCAGAAAAATGT 1560
Db 1608 TTGGTTAAGCAGCAGATTTTAAATATGACTACCTTTGACCACCAAGAACTCAGAAAAATGT 1667
Qy 1561 GAAACTTTTTCAGTCCCTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCACG 1620
Db 1668 GAAACTTTTTCAGTCCCTTCTCAGGAAGTTCTGATGGGACAATCTTATAGTCCACTCCAG 1727
Qy 1621 GCCACGGCAAAAGACTACAGATGGCTTAATGGGGTGCACCTTGCACATCAAAACT 1680
Db 1728 GCAGCCGCAAAAGAGCCTCAAGTGTGTCTTAATGGGGTCTCCAGTTTGGCATGTCTAAACT 1787
Qy 1681 GACTAAATCTCTCTCGCTCCACCTTCTACATTCAGACTTTGCCAGACACATTCATGTGT 1740
Db 1788 TACTAAATCTCTCTCGCTCCACCTTCCACTTCAGACTTTGCCAGACACGTTCTCGCAT 1847
Qy 1741 GTCTGAACACAGTTCCATCAGTGTGCTGAATATATACTCTCGAAGAAAGTAAACCAAGTGA 1800
Db 1848 ATCTGAACATAGTTCAATCAATGACTGCTGAATATACTGCTGAAGAAATTAACCAAAATCA 1907
Qy 1801 GGTGCAAGAAAGCAGGATCAGAGTGCAGTGCAGTGCAGGCGCCAGCTCCCGGA 1860
Db 1908 GGTGGAGTAGAATGTACAAAATCAAAAATGAGTGTGGCATCAAGACCTGGATCACAGGA 1967
Qy 1861 GGGGTGCTACAGCGGATGCTCTCGGCTTTCAGAGCGCTCACGGGACCGAGATGACTT 1920
Db 1968 AGTGTCTATAGTGGATGCTCTTGGAGTACACAAATTCATGCTAGTAAAGAGATGACTT 2027
Qy 1921 ACCTTAAATGTCGGGCTGCAGTGTCTGTTCCAGATGTCGCTAGAGG-----A 1969
Db 2028 ACCTTAGACATGTGGACTGGAATTTTTCATTAATGTTGTTTCATCAAGTTTTCACATCTAA 2087
Qy 1970 GTTACACAGGGTGTAGCATAAAGTCACTGCTC-----TAACTTAAGATGCTCAGAGTTG 2024
Db 2088 GTTGAACAGGGTGTGTCAAGATCACTGTTATCTTAATAACTTAAAGATGCTCTGAGTTG 2147
Qy 2025 TTTGTTTGGACTTCGCTCTCTTCCCAAGAGCTGAAATGCT-AAGCTACTTAAAGGA 2083
Db 2148 TTTGTTTGGACTTCCTGCTCTTCCCAAGAGTGAATCTTAAATCTTAAATCTTAAAGGA 2207
Qy 2084 TGCAAAAGCTTTGG-----TTGTGTGTAGTAAACAGAGCCCTGCTCTGTGACTGCGAGGA 2139
Db 2208 TATAAAGCTTTGGATATGATTTTGTAGTAAACAGAGCATCTGCTTCTGTGAATAAAGGA 2267
Qy 2140 ATGCAT-GGCGTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCGAGTACCG 2196
Db 2268 ATGTATAGATGTTGGATGGAACAGAGCGTACAGTGTCTTCTTTATAGGTATTA 2327
Qy 2197 AGAGAGCACTTTTAGGGA 2215
Db 2328 AAATAGCACTTTTAGGAA 2346

RESULT 12
LOCUS BC064389
DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (CDNA clone MGC:75026 IMAGE:5745142), complete cds.
ACCESSION BC064389
VERSION BC064389.1 GI:39963669
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2540)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
```

```
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Scapleton,L., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalski,U., Smalusi,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2540)
Director MGC Project.
Direct Submission
Submitted (15-DEC-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgaps-x@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc\_mgc@hgrl.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
Maduro,Q.L., Masiello,C., Mapkari,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantropop,S., Thomas,P.J., Touchman,J.W.,
Tsurgueon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAK Plate: 139 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 41281570.
FEATURES
Location/Qualifiers
1..2540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:75026 IMAGE:5745142"
/tissue_type="Brain, adult medulla"
/clone_lib="NIH MGC_119"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..2540
/genes="SSX2IP"
/notes="synonyms: KIAA0923, ADIP"
/db_xref="GeneID:117178"
/db_xref="MIM:608690"
241..1920
/genes="SSX2IP"
/codon_start=1
/product="SSX2IP protein"
/protein_id="AAH64389.1"
gene
CDS
```



/db\_xref="GI:39963670"  
/db\_xref="GeneID:117178"  
/db\_xref="MIM:608690"  
/translation="MGDMVTDPGLSSSEKTSIQVTSKMPSSLSXQQVLCSSIP  
LSKNVHSFSAFTCTEDNIISQISYLDQELTTFGFPSPSYBESKGTRELNIVAVLNC  
MNELLVORNLKLLAENKEDVOKLONLIIASRATQYNHMKREYENKLERLHQLVMN  
OLOCKNRNLKLLAENKEDVOKLONLIIASRATQYNHMKREYENKLERLHQLVMN  
KDKKIAMDILNYVGRADGKRSWGTGTEARNEDEMYKILLNDYEQKQILMENAE  
LKVLQOMKEMISLSPOKKPRERVDSTGTIVSDVEEDAGELSRSMWDLSCETV  
RQLNIRKQWRILLKSHVEKLDNQVSKVHLEFNDEVDISQDHEQTEKLELEIQ  
CHEMTKQOQLQALATAYDDTTSLLRDLYLLEBKERLKEWSLFKQKNFERERR  
SFTEAARILGLERKAFEEERASWLKQOFLNMTTFDHOENSENVKLFSAFGSSDMDNLI  
VHSRQPKPHSVNSGSPVMSKLTSLPASPSSTDFTQTRSCISEHRAKN"

ORIGIN

Query Match 47.9%; Score 1290.6; DB 5; Length 2540;  
Best Local Similarity 83.1%; Pred. No. 0;  
Matches 1496; Conservative 0; Mismatches 299; Indels 6; Gaps 2;  
QY 1 CGTAGGAGTACAGGAGCTGTTGTAAAGCGTCGAGCACTGAGCGGCTCTCTAGGTAT 60  
DB 162 CATAGAAGAGTGACAGCAGCTAGACTAAATGTTAACTGCTGAACCTAGTCTCAGGTAT 221  
QY 61 CCGTGGCTCTGGAACCTTGCTATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC 120  
DB 222 CCGTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTGACAGATCCAGGTCTGTCTC 281  
QY 121 AGAAAAAATACTCTCTCAATATACCTCAGAAAAAAGATGTCTCCGTCAGTTTGTGA 180  
DB 282 AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTTATA 341  
QY 181 CTCACGACGAGTCTGTGCTCTTCTAGTACCTTTATCCAAAAACGTGCGATGGTGTTCGG 240  
DB 342 CTCACGACGAGTCTGTGCTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG 401  
QY 241 TGTCTTCTGCACGAGAGCAATTTGAACAAAGTATTTCTATCTTGTATCGAGCTGAC 300  
DB 402 TGCCTTCTGCACAGAAGATATATTGAACAGAGTATCTCATATCTTGTATCAGAAATTGAC 461  
QY 301 CACCTTCGGGTTTCCTTCCTGTATGAAGAAATCCAAAAAGTAAAGAGGCAAAAGAGAGAATT 360  
DB 462 TACTTTTGGTTTTCTTCATTATATGAAGAAATCCAAAGTAAAGAGACAAAGAGAGATT 521  
QY 361 AAATATAGTCGTGTTCTGAACTGTATGAACGAGTCTGCTGCTTTCAGCGGAAGAACT 420  
DB 522 AAATATAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581  
QY 421 GCTGCCCGAGGAGCGTGGACACAGAACTTTGAAGCTGGGACGTGACATGGACCACT 480  
DB 582 TCTAGCTCAGGAAATGTGGAGACAGAAATTTGAAGCTGGGAGTGAATGGACCATCT 641  
QY 481 GCAGAGCTGCTACGCGCAAACTTTAAGGAGCAGTTGGAAACGCTCCAGCGGGAGATGATCGG 540  
DB 642 ACAGAGCTGCTACTCAAAACTTTAAGGAACTTGAAGAACTCCAGGAGGGAATGATTTGG 701  
QY 541 GCTTCAAGAGAGACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
DB 702 GCTTCAGGAAAGAGACAGACAGTTCACAAATGTAAGAACAGGAATTTGTCATCAGCTACTAAA 761  
QY 601 GAAATCAGAAAGATGAGGTACAAAATTTACAAAATATCAATTAAGCTTAAAGGAGCGCTGCATCA 720  
DB 762 GAATGAGAAAGATGAGGTGCAAAAATTTACAAAATATCAATTAAGCTTAAAGGAGCGCTGCATCA 821  
QY 661 TAATCATGATGTGAAGAGAGAGCGGTGAATATTAATTAAGCTTAAAGGAGCGCTGCATCA 720  
DB 822 TAATCATGATGTGAAGAGAGAGCGGTGAATATTAATTAAGCTTAAAGGAGCGCTGCATCA 881  
QY 721 GCTCGTTATGAACAAGAGAGTAAAAATAGCCATGGATGTTTAAATATGTGGGTCG 780  
DB 882 ACTTGTATGAACAAGAGAGTAAAAATAGCTATGGACATTTTGAATTTATGTGGGAG 941  
QY 781 AGCTGATCGCAACGAGCTCATGAGGAGCTGACAAAAACAGAGCCAGGAATGAAGATGA 840

DB 942 ACCTGATGGAAAAAGAGGCTCTCGAGGACTCGTGAATAAAGCTGAAGCCAGGAATGAAGATGA 1001  
QY 841 GATGTACAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAAAGCAGATCCTGATGGAGAA 900  
DB 1002 AATGTATAAAAATTTCTTTGAATGATTATGAATATATCGTCAGAAACAAATCCTAATGGAAA 1061  
QY 901 CGCGAGCTGAAGAAGGCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCTCC 960  
DB 1062 TGCAGAACTTAAGAAAGGTTCTTCAACAAATGAAGAAAGGAATGATTTCTCTCTCTCTCC 1121  
QY 961 TCAGAAAGAAAGCCAGCGGAAAGAGCAGAGGACGCGACAGCGACCTGTTGCTATCTCCGA 1020  
DB 1122 CCAAAAGAAAGAACTTAGAGAAAGAGTAGTAGTACAGGAACCTGT---TATTCCGA 1178  
QY 1021 TATGAAGATGACTCTCGGGAACTCGAGCAGAGACAGCGTGTGGGGCTTTTCTCTGTGACAC 1080  
DB 1179 TGTTCAGAGAGATGCGCGGGAACCTAAGCAGAGAGATGATGTGGGACCTTTTCTGTGAAAC 1238  
QY 1081 TGTGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGATCATGT 1140  
DB 1239 TGTGAGAGCAGCTTTACAAAACAGCATCAGAAACAGTGGAGAAATTTTGAAGATCATGT 1298  
QY 1141 AGAAAACTTCGATAACCAAGCTTTCGAAGGTACACTCAGAGGSCCTTAATGAGGAGACGT 1200  
DB 1299 AGAAAGCTTGATAACCAAGTTTCAAGGTACACTCGGAAGSTTTTAATGATGAAGATGT 1358  
QY 1201 CATCTCAGCAAGACCATGAGCAAGAGACTGAGAACTTGGAGCTGGAGATTTGAGCGGTG 1260  
DB 1359 AATCTCAGCAAGACCATGAACAGAACTGAAAACTCGAGTTAGAAATTTGACAGTG 1418  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGGCCACACGCTGTA 1320  
DB 1419 TAAAGAAATGATTAATACTCAGCAACAGCTTTTA---CAGCAGCTCGCTACTGCAATGA 1475  
QY 1321 TGATGACACCACTCAGCTGTTGCGAGAGCTGTACTTGTGGAAGAAAAAGGACCGCTTAA 1380  
DB 1476 TGATGATACCACTTCACTATTAACGAGAGCTGTATTTGTTGGAAGAAAAAGGACCGCTCAA 1535  
QY 1381 AGAAGAGTGGACCCCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGCGCAAGCTT 1440  
DB 1536 AGAAGATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGAGAGACGAAGCTT 1595  
QY 1441 TACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCCAG 1500  
DB 1596 TACAGAGCGCTTATTCGCTCGGATTTGGAGAGAAAGGCAATTTGAAGAGAAAGAGCCAG 1655  
QY 1501 CTGGTAAAGCAGCAGTTTTTAAACATGACCGAACTTTGACCCACCAAGAACTCAGAAAAATGT 1560  
DB 1656 TTGTTAAAGCAGCAGTTTTCTAAATATAGCTACCTTTGACCCACCAAGAACTCAGAAAAATGT 1715  
QY 1561 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAG 1620  
DB 1716 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTTGGGACAATCTTATAGTGCCTCAG 1775  
QY 1621 GCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCTTCGCATCAAAACT 1680  
DB 1776 GCAGCGCAAAAGAGCGCTCAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTTAAACT 1835  
QY 1681 GACTAAATCT 1740  
DB 1836 TACTAAATCT 1895  
QY 1741 GTCTGAACACAGTTCATCAGTGTGCTGAATATAAATCTCCTGAAGAAAGTAAACCAAGTGA 1800  
DB 1896 ATCTGAACATAGGCAAGAACTCAGAAAAAGCCTGTTCTCTGTTGAGTGAAGAAAGATTTGT 1955  
QY 1801 G 1801  
DB 1956 G 1956

LOCUS BD156617 2716 bp DNA linear PAT 17-JAN-2003  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.

ACCESSION BD156617  
VERSION BD156617.1 GI:27862375  
KEYWORDS JP 2002191363-A/11460.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 2716)  
Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,  
Ishii, S., Sudiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.  
Primer for synthesizing full-length cDNA and use thereof  
Patent: JP 2002191363-A 11460 09-JUL-2002;  
HELIX RESEARCH INSTITUTE

COMMENT OS Homo sapiens (human)  
PN JP 2002191363-A/11460  
PD 09-JUL-2002  
PF 28-JUL-2000 JP 200280990  
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU  
SAITO,  
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,  
PI KEIICHI NAGAI, TETSUJI OTSUKI  
PC

C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
Location/Qualifiers (78) (1328).

FT CDS Location/Qualifiers  
source 1..2716  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 46.8%; Score 1260.4; DB 2; Length 2716;  
Best Local Similarity 79.7%; Pred. No. 0;  
Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;  
QY 594 TCCTCAACAATGAGAAAGATGAGGTACAAATAATACAAAATATCATAGCCAGCGGGCTA 653  
DB 1 TACTAAAGAATGAGAAAGATGAGGTGCAAAATATACAAAATATCATTCGAAGTCAAGCTA 60  
QY 654 CTCAGTAAATCATGATGTGAAGAGGAAGAGCGTGAATATAATAAGCTAAAGAGCGCC 713  
DB 61 CTCAGTAAATCATGATATGAAGAGAAAGAGCGTGAATATAATAAAGCTGAAGGAACGTC 120  
QY 714 TGCATCAGCTCGTTATGAACAAGAGGATATAAATCATAGCCATGATGTTTAAATATG 773  
DB 121 TACATCAACTTGTGTATGAACAAGAGGATGAAGAAATAGCTATGGACATTTTGAATTATG 180  
QY 774 TGGGTCCAGCTGTGGCAACGAGCTCATCGAGGACTGACAAACAGCAAGCAGGGAATG 833  
DB 181 TCGGAGAGCTGTATGAAGAAAGAGGCTCTCGAGGACTGGTAAATCTGAGCCAGGAATG 240  
QY 834 AAGATGAGATGTACAAAATCTCTTGAATGATTATGATACCGCCAGAGCAGATCCTGA 893  
DB 241 AAGATGAATGTATAAAATCTCTTGAATGATTATGATATCGTCAGAAAACAAATCCTAA 300  
QY 894 TGGGAACCGGAGCTGAAGAAGTCTCTCAGCAGATGAAGAAGGAGATGATCTCTCC 953  
DB 301 TGGAAAATGCGAGAACTTAAGAAGGTCTTCAACAAATGAAAAGGAAATGATTTCTCTTC 360  
QY 954 TGTCTCCTCAGAGAAGAGCCCGGAAAGAGCAGAGGAGCGGACAGGCACTGTTGGCTA 1013  
DB 361 TTTCTCCCCAAAGAAAGAAACCTAGAGAAAGATGATAGTACAGAACTGT--TA 417  
QY 1014 TCTCCGATATAGAAGATCACTCTGGGAACTGAGCAGAGACGCGTGTGGGCGCTTTCCT 1073  
DB 418 TTTCCGATGTTGAAGAAGATGCGGGAACTTAAGCAGAGAGAGTATGTGGGACCTTTCCT 477

QY 1074 GTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGAAACAGTGGAGAAATTTGAAAA 1133  
DB 478 GTGAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTGAAA 537  
QY 1134 GTCATGTAGAAAAAATCGATAAACCAAGCTTCGAAAGTACACCTGGAAAGTTTAAATGATG 1193  
DB 538 GTCATGTAGAAAGCTGTGATAACCAAGTTTCAAAGGTACACCTGGAAAGTTTAAATGATG 597  
QY 1194 AGGACGTCATCTCAGCAAGAGCATGAGCAAGAGACTGAGAAACTGAGAGCTGAGATTG 1253  
DB 598 AAGATGTAATCTCAGCAAGAGCATGAGCAAGAAACTGAAAACTCAGTTAGAAATTC 657  
QY 1254 AGCGGTGAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCTGCCACCA 1313  
DB 658 AGCAGTGTAAAGAAATGATTAAACCTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTG 717  
QY 1314 CGTGTGATGATGACACCACCTCCTACTGTTGGGAGACTGTTACTTGTGGAAGAAAGAAC 1373  
DB 718 CATATGATGATGATACCACTTCACATTACGAGACTGTTATTGTTGGAAGAAAGAAC 777  
QY 1374 GCCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAGAAATTTTGAGAGAGAAAGGC 1433  
DB 778 GTCTCAAAGAGAAATGCTGCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGAGAGAC 837  
QY 1434 GAAGCTTTACAGAGCTGCCATTCGATTGGGTTGGAGAGAAAGGCGTTTGAAGAAAGAGC 1493  
DB 838 GAAGCTTTACAGAGCGCGCTATTTCGCCCTGGGATTTGGAGAGATGGCAATTTGGAAGAGAA 897  
QY 1494 GAGCCAGCTGGGTAAGCAGCAGTGTAAACATGACGAACTTTGACCAACAGAACTCAG 1553  
DB 898 GAGCCAGTGGTTAAAGCAGCAGTGTCTAAATATGACTACTCTTTGACCAACAGAACTCAG 957  
QY 1554 AAAATGTGAAACTTTTCAGTGCCTTCTCAGAGAGTCTTGATCAGAGAACTTTATAGTCC 1613  
DB 958 AAAATGTGAAACTTTTCAGTGCCTTCTCAGAGAAATTTCTGATTTGGGACAACTTTATAGTGC 1017  
QY 1614 ACTCAGCGCCACGCAAAAGAGCTACACAGTGTGGCTAAATGGGTTGCCAGTTCACAT 1673  
DB 1018 ACTCAGGCGCAGCGCAAAAGAGCTCAGAGTGTCTAATGGTCTCCAGTTCGATGT 1077  
QY 1674 CAAAACCTGACTAAATCTCTTCGCCCTCACCTTCTACTTCTCAGACTTTGCCAGACACATT 1733  
DB 1078 CTAACCTTACTAAATCTCTTCCTGCTTCCACTTCCACTTCCAGACTTTTGCAGACAGCTT 1137  
QY 1734 CATGTGTCTGAACACAGTTCATCAGTGTCTGAATATAAATCTCCTGAGAGAAAGTAAAC 1793  
DB 1138 CCTGCATATCTGAACATAGTTCAATCAATGTACTGAATATAAATGCTGGAAGAAATTAAC 1197  
QY 1794 CAAGTGAAGTTTGAAGAGAAAGCAGCGATCAGAAAGTGGAGCGTGCAGTCCAGGCGCCAGCT 1853  
DB 1198 CAAATCAGGTTGGAGGAGATGTACAATCAAAATGGAGTGTGGCGTCAAGACTGGAT 1257  
QY 1854 CGCGGAGGGGTGCTACAGCGGATGCTCCTCGGCCCTTTCAGAGCGCTCACGGGAGCCGAG 1913  
DB 1258 CACAGGAAGGTGCTATAGTGGATGCTCCTTGAGCTACACAAATCTCATGTAGAAAAAG 1317  
QY 1914 ATGACTTACCTTAAATGTGGGGTGCAGTGTCTTCCAGATGTGGCTAGAGG----- 1968  
DB 1318 ATGACTTACCTTACATGTGGACTGGAAATTTTTCATTAATGTGTTCATCAAGTTTCA 1377  
QY 1969 -----AGTTTGACACAGGGGTGTAGCATAAAAGTCAAGTGGTC-----TAACCTTAAGATGCTC 2017  
DB 1378 CATCTAAGTTGAAACAGGGTGTGTCATAAAGTCAAGTCACTTCTCTAATAACTTAAGATGCTC 1437  
QY 2018 AGAGTGTGTTGTTGGACTTGGCTGTCTTCCCCCAAGAGAGCTGAAATGCT--AAGCTACTT 2076  
DB 1438 TGAGTGTGTTGTTGGACTTCCCTGTCTTCCCCCAAGAGAGTGAATCTTAAATCTATTT 1497  
QY 2077 AAAAGGATGCAAGACTTTGG-----TTGTGTGTTAGTAAACAGAGCCCTGCTCTGTGAC 2132  
DB 1498 AAAAGGATATAAAAGACTTTGGATATGATTTTATAGTAACAGAGCATCTGGTCTGTGAA 1557

```
QY 2133 TGCAGGAATGCAT-GGCGTTTGATGGAACAGAGCGCTGGAATGA--TTGCCTCGCA 2189
Db 1558 TAAAGGATGTAGATGTTTGATGGAACAAAGCACTAGACGAGTTCCTCTTATA 1617
QY 2190 GGTACCGAAGAGACACATTTTAGGAGCTGTTCTGTAAACATTAATAATTCGTCCTCAAG 2249
Db 1618 GGTATTAATAATAGCACATTTTAGGAAACTGATTATTTGTAATGTTAAATTTGTCCTCAA 1677
QY 2250 TGTGTTGTCATGGAAGTGTAGCCTTTACTTGAATGTATATCTAGATTTTAAACAAG 2309
Db 1678 TATAGTTGTCATGGAAGTGTAGCCTTTACTTGAATGTATATCTAGATTTTAAACAAG 1737
QY 2310 CAGGTTCTATATTTATTTAGTTAGTGATTTTGGGATTACCTCTTTCATATGTTTT-- 2367
Db 1738 CGAGTTCTATATTTATTTAGTTAGTG-GTTTGAATTAACCTCTTTCATATGTTTTAA 1796
QY 2368 -----GTCTCTGTACATAAATATACATGACTATGTTAAGAGGCTTT 2408
Db 1797 ATAAAGTGAAATTTATGTATGTTTGTACATAGATACATGATTATGTTAAGAGGCTTT 1856
QY 2409 AAGGTTTAAAACTTCACACCATGCTTGAGTATAGCATTTTCAGCAATTAATAATGTTTT 2468
Db 1857 AAGATTTAAAGTTTCACA-CAACCATAGTAGTATTTTCATGCC-AGTAAATTTTTT 1914
QY 2469 CAGTGGCATGTTTACAGAGG--TTAGGACCATGCCACATGACAGTTAAGACTTTAT 2526
Db 1915 TAGTGGTATCTGTTTACAGATGTATTAGGACCATGTATGCAATACATTTAAGA-ATTCT 1973
QY 2527 TTTTAAAGCCATCGGCAATAAAATTCAAAGCCCTTCAT-AAGCTGAGTTC---AGAT 2582
Db 1974 CTTTAAATACATCGGCAATAAATTTGAAAGTATTCAGAACTGAGTTCCTTAGAT 2033
QY 2583 AACTAGACTACTAAGTTTACATTTTGTAGATTTTAAAGCAAT-GTATTTTATTTATA 2641
Db 2034 AATCAACACTACTAACAATTTACATTTTGTAGATTTTATGACATTAAGATTTTATTTGTA 2093
QY 2642 TATGT-GAATGTATAATTTCTAAGAGGAATATTCATTTAGTATGAGTAATGGG 2692
Db 2094 TATGTAGATATTATAATTTTTTAAAGGACTATTGATGATAGAATAAGG 2145

RESULT 14
AX877359 LOCUS 2716 bp DNA linear PAT 17-DEC-2003
DEFINITION Sequence 12264 from Patent EP1074617.
ACCESSION AX877359
VERSION 1
KEYWORDS AX877359.1 GI:40032095
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
REFERENCE
AUTHORS Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
TITLE Primers for synthesising full-length cDNA and their use
JOURNAL Patent: EP 1074617-A 12264 07-FEB-2001;
Research Association for Biotechnology (JP)
FEATURES
source
1. 2716
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
78..1331
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE89690.1"
/db_xref="GI:40032096"
/translation="MKREREYNKLERLHOLVNMKKDKIAMDILNVGRADKRG
WRGKTEARNDEWYKILNDYERQKILMENAELKKVLQOMKEMISLLSPKKKP
RERVDDSTGTVISVEDEAGELSRSMWDLSCETVRSLTNSIKQWRLKSHVEKLD
NQSVKHLGEGNDEVDVSRQHEQTEKLELEIQOCKEMIKTQOQLQOQLATAYDD
```

## ORIGIN

```
Query Match 46.8%; Score 1260.4; DB 2; Length 2716;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;

QY 594 TCCTCAAGAATCAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCAGCGGCTA 653
Db 1 TACTAAAGATGAGAAAGATGAGGTGCAAAATTTACAAAATATCATTTGCAAGTCAGCTA 60
QY 654 CTCAGTATAATCATGATGTGAAGAGGAGCGGTGAATATAATAAAGCTAAAGAGCGCC 713
Db 61 CTCAGTATAATCATGATGTGAAGAGGAGCGGTGAATATAATAACTGAAGGAACGTC 120
QY 714 TGCATCAGCTCGTTATGAACAAGAGGATAAACATAGCCATGGATGTTTTAAATATG 773
Db 121 TACATCAAATCTGTTATGAACAAGAGGATAAAGAAATAGCTATGGACATTTTGAATATG 180
QY 774 TGGGTGAGCTGTGCGCAACAGGCTCATCGAGGACTGACAAAACAGAACGACGAAATG 833
Db 181 TCGGAGAGCTGTGGAAGAGGCTCTCGAGGACTGGTAAACTGAAGCCAGAAATG 240
QY 834 AAGATGAGATGTACAAAATTTCTGTTGAATGATTATGATACCGCCAGAGCAGATCTTGA 893
Db 241 AAGATGAATGTATAAATTTCTTTGATGATTATGATATCGTCAGAAAACAAATCTTAA 300
QY 894 TGGAGAACCGGAGCTGAAAGAGGTCCTCAGCAGATGAAGAAGAGATGATCTCTCTCC 953
Db 301 TGGAAAATGCAAGAACTTAAAGAGGTTCTTCAACAAATGAAAAGGAAATGATTCTCTTC 360
QY 954 TGTCTCTCACAAGAAGAGCCAGGGAAGAGCAGAGAGCGGACAGGCACTGTTGCTA 1013
Db 361 TTTCTCCCAAGAGAAACCTAGAGAAAGATAGTAGTAGTACAGAACTGT---TA 417
QY 1014 TCTCCGATATAGAAGTGAATCTCGGGAACTGAGCAGAGACAGCGGTGGGGCTTTCT 1073
Db 418 TTTCCGATGTTGAAGAGATCGCGGAACTAAGCAGAGAGATGATGCGGACCTTTCTCT 477
QY 1074 GTGACATGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGAGAAATTTGAAA 1133
Db 478 GTGAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAAAACAGTGGAGAAATTTGAAA 537
QY 1134 GTCATGTAGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGG 1193
Db 538 GTCATGTAGAAAAAGCTTGTATACCAAGTTTCAAAGGTACACCTGGAAGGTTTTAATGATG 597
QY 1194 AGGAGCTCATCTCAGCAAGACCATGAGCAAGAGACTGAGAAAACCTGAGGCTGGAGATTG 1253
Db 598 AAGATGTAATCTCAGCAAGACCATGACAAAGAACTGAAAACCTCGAGTTAGAAATTC 657
QY 1254 AGCGGTGTAAGAGATGATCAAGGCTCAGCAGCAGCTCTTTACAGCAGCAGCTGGCCACA 1313
Db 658 AGCAGTGTAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGTACTG 717
QY 1314 CGTGTGATGATGACACACCTCAGCTGTTGGAGACTGTTACTCTGCTGAAGAAAGGAAC 1373
Db 718 CATATGATGATGATACCACTTCACTATTACAGACTGTTATTGTTGGAAGAAAGGAAC 777
QY 1374 GCCTTAAAGAGAGTGGAGCCCTTTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGC 1433
Db 778 GTCTCAAAGAGAAATGTCCTTTTTTAAAGAGCAGAAAAGAAATTTTGAAGAGGAGAGAC 837
QY 1434 GAAGCTTTACAGAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGGTTTGAAGAAGAGC 1493
Db 838 GAAGCTTTACAGAGCGCGCTATTTCGCTGGGATTTGGAGAGAAATGGCATTTTGAAGAAGAA 897
QY 1494 GAGCCAGCTGGTAAAGCAGCAGTTTTTAAACATGACGAACTTTTGAACCCAGACTCAG 1553
Db 898 GAGCCAGTTGGTTAAAGCAGCAGTTTTCTTAAATATGACTTACCTTTTGACCCAGAACTCAG 957
```

```
TTSLRDYLLBEKERLKEBSLKFQKKNFERRRSTFAAIRLGLRMAFEERRAS
WKQOFLNMTTFDHQNSNVLFSFSGSDNDNLIVHSRQPKPHSVNSGSPVCM
KLTKSLPASPTSDFCQTRCSIFSHSSINVLITAEIKPNQVGECTNQKWSVASRP
GSQEGCYSGCSLSYNSHVEKDDLDP"
```

Qy 1554 AAAATGTGAACCTTTTCAGTGCCTCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCC 1613  
Db 958 AAAATGTGAACCTTTTCAGTGCCTCTCTCAGGAAGTTCTGATGGACAATCTTATAGTGC 1017  
Qy 1614 ACTCAGGCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCACGCTTGCACAT 1673  
Db 1018 ACTCGAGCGCCGCAAAAGAGCTCACAGTGTCTTAATGGGTCTCCAGTTTGCATGT 1077  
Qy 1674 CAAAACCTGAATAATCTCTCTGCTGCCTCACCTTCTATCTTCAGACTTTTGGCCAGACATTT 1733  
Db 1078 CTAAACCTTACTAAATCTCTCTGCTGCCTCACCTTCCACTTTCAGACTTTTGGCCAGACAGTT 1137  
Qy 1734 CATGTGCTCTGAACACAGTTCCTCATCTAGTGTCTGAATATACTTAACTCCTGGAAGAAAGTAAAC 1793  
Db 1138 CCTGCATATCTGAACATAGTTCAATCAATGACTGAATATACTTAACTGCTGAAGAAATTAAC 1197  
Qy 1794 CAAGTGAGGTTGCAAGAGAAAGCGGATCAGAGTGGAGGTGAGCTGAGTCGAGGCCACGCT 1853  
Db 1198 CAATCAGGTTGGAGGAGATGTACAAATCAAAATGAGGTGGCGTCAAGACTGGAT 1257  
Qy 1854 CGCGGAGGGTGTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCGAG 1913  
Db 1258 CACAGGAAGGTTGCTATAGTGGATGCTCTCTTGAGCTACACAAATCTCATGTAGAAAAG 1317  
Qy 1914 ATGACTTACCTTAATGTGCGGGTGCAGTCTGTTCCAGATGCGCTAGAGG---- 1968  
Db 1318 ATGACTTACCTTAGACATGTGGAGTGGAAATTTTTTTTCAATTAATGTGTTCACTCAAGTTTCA 1377  
Qy 1969 -----AGTGACACAGGTTGTAGCATAAAGTCAGTCTGTC-----TAACCTTAAGATGCTC 2017  
Db 1378 CATCTAAGTTGAAACAGGTTGTGCATAAAGTCAGTTATCTCTAATAACTTAAGATGGTC 1437  
Qy 2018 AGAGTTGTTTGTGGACTTGCCTGCTCTTCCGCCAAAGAGCTGAAATGCT--AAGTACTTT 2076  
Db 1438 TGAGTTGTTTGTGGACTTGCCTGCTCTTCCGCCAAAGAGTTGAAATCTTAATCTATTT 1497  
Qy 2077 AAAAGGATGCAAGCTTTGG----TTGTGCTGTAGTACAGAGCCCTGCTGCTGAC 2132  
Db 1498 AAAAGGATATAAAGCTTTGGATATGTATTTTTTTAGTAAACAGAGCATCTGGTCTGTGAA 1557  
Qy 2133 TGCAGGAATGATC--GGCTTTTGGATGGAAACAGAGCCCTGGAATGA--TTGCCCTCGCCA 2189  
Db 1558 TAAAGGATGTATAGATGTTTGGATGGAAACAAAGACACTAGACTGAGTTTCTCTTTATA 1617  
Qy 2190 GGTACCGAGAGAGCACTTTTAGGACTGGTTTCTGTAAACATTAATATTCGTTCCCAAG 2249  
Db 1618 GGTATTAATAAATAGCACTTTTAGGAACTGATTAATGTAATGTTTAAATTTTGTCTCAA 1677  
Qy 2250 TGTGTTGGCATTTGGAAGTGTAGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 2309  
Db 1678 TATAGTTGGCATTTGGAAGTTTAGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 1737  
Qy 2310 CAGGTTCTATATTATTTATGTTAGTGTGATTTTGGGATTAACCTCTTTTCATATGTTTT-- 2367  
Db 1738 CGAGTTCTATATTATTTATGTTAGTGTG--GTTTGAATTAACCTCTTTTCATATGTTTTAA 1796  
Qy 2368 -----GTGCTGTACATAAATATACATGACTATGTTTAAGAGGCTTT 2408  
Db 1797 ATAAGTGAAATTTATGTATGTTTGTATACATAGATACATGATTTATGTTTAAGAGGCTTT 1856  
Qy 2409 AAGGTTTAAAACTTCACACCATGCTTCAGTATAGCATTTTCATGCCAATTAATAATGTTTT 2468  
Db 1857 AAGATTTAAAGTTTCACA--CAACCATTAAGTATAGTATTTTCATGCC--AGTAAATTTTTT 1914  
Qy 2469 CAGTGGCATGTTGTTTACAGAGG--TTAGGACCACTGCCACATGACAGTAAAGACTTTAT 2526  
Db 1915 TAGTGGTATTCTGTTTACAGATGTTATAGGACCATTTGATGCTATACATTTAAGA--ATTCT 1973  
Qy 2527 TTTTAAAGCATCTGGGCAATAAATTCAAAGCCCTTCAT--AAGCTGAGTTC---AGAT 2582  
Db 1974 CTTTAAATACATCTGGGCAATAAATTTGAAAGGTTATTCCTGAAGAGCTGAGTTCTTTAGAT 2033

Qy 2583 AACTAGAACTACTAAAGCTTACTATTTTGTGAGATTTTAAAGCAAT--GTATTTTATTTTATA 2641  
Db 2034 AATCAACACTACTAAACATTACATTTTGTGAGATTTTATGACATTAGATTTTATTTTCTA 2093  
Qy 2642 TATGT--GAATGTATATAATTTCTTAAGAGGAATATGATTATGGAGTAATGGG 2692  
Db 2094 TATGTAGAATATTATATAATTTTAAAGGACTTATGATGATAGAGAATAGG 2145  
RESULT 15  
AK001710 2716 bp mRNA linear PRI 20-JAN-2006  
LOCUS Homo sapiens cDNA FLJ10848 fis, clone NT2RP4001407, highly similar  
to Homo sapiens mRNA for KIAA0923 protein.  
AK001710  
AK001710.1 GI:7023139  
VERSION oligo capping; fis (full insert sequence).  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
REFERENCE  
1  
Oca,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R.,  
Wakamatsu,A., Hayaashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H.,  
Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y.,  
Nagahara,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuna,M.,  
Shiratori,A., Sudo,H., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,  
Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T.,  
Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K.,  
Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H.,  
Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M.,  
Hirooka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S.,  
Yosida,M., Hotta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A.,  
Hara,H., Tanase,T.O., Nomura,Y., Togiya,S., Komai,F., Hara,R.,  
Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A.,  
Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichiwara,T.,  
Shiobata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S.,  
Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H.,  
Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T.,  
Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K.,  
Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M.,  
Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y.,  
Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiji,T., Kobatake,N.,  
Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T.,  
Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K.,  
Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T.,  
Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J.,  
Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K.,  
Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R.,  
Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.  
Complete sequencing and characterization of 21,243 full-length  
human cDNAs  
Nat. Genet. 36 (1), 40-45 (2004)  
1470239  
2  
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,  
Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,  
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,  
Masuho,Y. and Kanehori,K.  
NEDO human cDNA sequencing project  
Unpublished  
3 (bases 1 to 2716)  
Isogai,T. and Otsuki,T.  
Direct Submission  
Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,  
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
(E-mail: flj-cdna@nifty.com, Tel.81-438-52-3975, Fax:81-438-52-3986)  
NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library

construction, 5'- & 3'-end one pass sequencing and clone selection:  
Helix Research Institute (supported by Japan Key Technology Center  
etc.) and Department of Virology, Institute of Medical Science,  
University of Tokyo.

## FEATURES

Location/Qualifiers

source

```

1..2716
/organism="Homo sapiens"
/mol_type="mrna"
/db_xref="taxon:9606"
/clone="NT2RP4001407"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP4"
/notes="cloning vector: pME18SFL3
mRNA from NT2 neuronal precursor cells after 2-weeks
retinoic acid (RA) induction."

```

## ORIGIN

```

Query Match          46.8%; Score 1260.4; DB 5; Length 2716;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1716; Conservative 0; Mismatches 376; Indels 60; Gaps 17;

QY 594 TCTGAAGAATGAGAAAGATGAGGTACAAAATTACAAAATATCATAGCCAGCCGGGCTA 653
DB 1 TACTAAAGAAATGAGAAAGATGAGGTGCAAAATTACAAAATATCATTCGAGTCGAGCTA 60
QY 654 CTCAGTAAATCATGATGTAAGAGGAAGGAGCGTGAATATATAAAGCTAAAGAGAGCC 713
DB 61 CTCAGTAAATCATGATGTAAGAGGAAGAGCGTGAATATAAATAACTGAAGGAACGTC 120
QY 714 TGCATCAGCTCGTTATGAACAGAGGATAAACATAGCCATGATGTTTAAATATG 773
DB 121 TACATCAACTTGTATGAACAGAGGAAGATAGAAAATAGCTATGACATTTTGAATATG 180
QY 774 TGGGTGAGCTGATGGCAAAAGAGGCTCATGGAGGACTGACAAAACAGAAAGCAGGAATG 833
DB 181 TCGGAGAGCTGATGGAAGAAAGAGGCTCTCGAGGACTGGTAAACTGAAGCCAGGAATG 240
QY 834 AAGATGAGATGTAACAAATCTCTGTTGAATGATTATGATACCGCAGAGAGATCTCTGA 893
DB 241 AAGATGAAATGTATAAAATCTCTTGAATGATTATGAATATCGTCAGAAAACAAATCTCTAA 300
QY 894 TGGAGAACCGGAGCTGAAGAGGTCTCCAGCAGATGAAGAGAGATGATCTCTCC 953
DB 301 TGGAAATGCGAGAACTTAAGAAAGGTCTTCAACAAATGAAAAGGAAATGATTTCTCTTC 360
QY 954 TGTCTCTCAGAAAGAAAGCCAGGGAAAGAGCAGAGAGCGGACAGGCACTGTTGCTA 1013
DB 361 TTTCTCCCAAAAGAGAAACCTAGAGAAAGTAGATAGTACAGGAATGTT--TA 417
QY 1014 TCTCGATATAGAGATGACTCTGGGAACTGAGCAGAGACAGCGTGTGGGCTTTTCT 1073
DB 418 TTTCCGATGTTGAAGAAAGATCCGGGAACTTAAGCAGAGAGATGATGTGGGACCTTCT 477
QY 1074 GTGACACTGTGAGAGAGAGCTGACAAACAGCATCAGAAACAGTGGAGAAATTTGAAA 1133
DB 478 GTGAACCTGTGAGAGAGAGCTTCAACACAGCATCAGAAAACAGTGGAGAAATTTGAAA 537
QY 1134 GTCATGTAGAAAAAATCGATACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGG 1193
DB 538 GTCATGTAGAAAAAGCTTGAATACCAAGTTTCAAGGTACACTCGAAGGTTTAAATGATG 597
QY 1194 AGGAGCTCATCTCAGCAAGAACCATGAGCAAGAGACTGAGAAAACCTGAGGCTGAGATTG 1253
DB 598 AAGATGTAATCTCAGCAAGAACCATGAGCAAGAACTGAAAACCTCGAGTTAGAAATTC 657
QY 1254 AGCGGTGAAGAGATGATCAGGCTCAGCAGAGCTCTTACAGCAGCAGCTGCCACCA 1313
DB 658 AGCAGTGTAAAGAAATGATTAATACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTG 717
QY 1314 CGTGTGATGATGACACCACTCACTGTTGCGAGACTGTTACTTCTCGGAAGAAAGGAAC 1373
DB 718 CATATGATGATGATACCACTTCACTATTTACAGACTGTTATTTGTTGGAAGAAAGGAAC 1856

```

```

QY 1374 GCCTTTAAAGAAAGAGTGGAGCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGC 1433
DB 778 GTCTCAAGAAAGAAATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGAGAGAC 837
QY 1434 GAAGCTTTACAGAAAGCTGCCATTCCGATTGGGGTTGGAGAGAAAGCGGTTTGAAGAGAGC 1493
DB 838 GAAGCTTTACAGAAAGCCGCTATTCCGCTGGGATGGAGAGAAATGGCAATTTGAAGAAGAAA 897
QY 1494 GAGCCAGCTGGGTAAAGCAGCAGATTTTAAACATGACCACTTTTACCAACCAAGAACTCAG 1553
DB 898 GAGCCAGTTGGTTAAAGCAGCAGATTTTAAATATGACTACCTTTGACCAACCAAGAACTCAG 957
QY 1554 AAAATGTGAATCTTTTTCAGTCCCTTCTCAGGAAGTTCTGATCCAGACAACTCTTATAGTCC 1613
DB 958 AAAATGTGAATCTTTTTCAGTCCCTTCTCAGGAAGTTCTGATGGGACAACTCTTATAGTGC 1017
QY 1614 ACTCAGCCGACGCAAGAAAGCTACACAGTGGCTTAATGGGTGCGAGCTTGCACAT 1673
DB 1018 ACTCAGGCGACGCGCAAGAAAGCTCAGAGTGTCTTAATGGGTCTCCAGTTTCATGT 1077
QY 1674 CAAAACCTGACTAAATCTCTCTGCTCACCTTCTACTTTCAGACTTTTCGCGCAGACACAT 1733
DB 1078 CTAAACTTACTTAATCTCTCTGCTCACCTTCCACTTCAGACTTTTCCAGACACGTT 1137
QY 1734 CATGTGTCTGAAACACAGATTTCCATCAGTGTCTGAATATAACTCTCTGAAAGAAAGTAAAC 1793
DB 1138 CCTGCATATCTGAACATAGTTCAATCAATGACTGAATATAACTGCTGAAGAAATTAAC 1197
QY 1794 CAAGTGAAGTTGCAAGAGAAAGCAGGATCAGAAAGTGGAGGTGCGAGTCGAGGCCAGCT 1853
DB 1198 CAAATCAGGTTGGAGGAGAAATGTACAAATCAAAAATGGAGTGTGGCGTCAAGACCTGGAT 1257
QY 1854 CGCGGAGGGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCCGAG 1913
DB 1258 CACAGGAAGTTGCTATAGTGGATGCTCTTGAGCTACAAATTTCTCATGTAGAAAAG 1317
QY 1914 ATGACTTACCTTAAATGTGGGGTGCAGTGTCTGTCCAGATGTGCGCTAGAGG- ---- 1968
DB 1318 ATGACTTACCTTAGACATGTGGAATGGAATTTTTCATTAATGTGTTCACTCAAGTTTCA 1377
QY 1969 -----AGTTGACACAGGTTGACATAAAGTCAAGTCTGTC- ----TAACTTAAAGTCTC 2017
DB 1378 CATCTAAGTTGAAAACAGGTTGTGTCATAAAGTCAAGTTATCTCTAATAACTTAAAGTGT 1437
QY 2018 AGAGTGTGTTGTTGGACTTGGCTGCTTCCGCCAAAGAGCTGAAATGCT--AAGTACTT 2076
DB 1438 TGAGTGTGTTGTTGGACTTCCCTGCTTTCCGCCAAAGAGTTGAATCTTAAATCTATTT 1497
QY 2077 AAAAGGATGCAAGCTTTGG- ----TTGTGTGTTAGTAAACAGAAAGCCCTTGGCTCTGTGAC 2132
DB 1498 AAAAGGATATAAAGCTTTGGATATGATTTTATAGTAACAGAGCATCTGGTTCTGTGAA 1557
QY 2133 TGCAGGAATGCAT--GGCGTTTGGATGGAACAGAGAGCGCTGGAATGA--TTGCTCTCGCA 2189
DB 1558 TAAAGGAATGTATAGATGTTTGGATGGAACAAAGACACTAGACTGAGTTTCTCTTATA 1617
QY 2190 GGTACCCAGAAAGAGCAGCTTTTAGGACTGTTCTGTAAACATTAATAATATTCGTCCCAAG 2249
DB 1618 GGTATTAATAATAGCAGCTTTTAGGAACTGATTAATGTTAAATGTTTATTTGTTCTCAA 1677
QY 2250 TGTGTTGGCATTGGAAGTGTAGCTTTTACTTGAATGTATATCTGTAGATTTTAAACAAAG 2309
DB 1678 TATAGTTGGCATTGGAAGTTTAGCTTTTACTTGAATGTATCTGTAGATTTTAAACAAAG 1737
QY 2310 CAGTTCTATATTTATTTATGTTAGTGTGATTTTGGGATTTACCTCTTTTCAATGTTTT-- 2367
DB 1738 CGAGTTCTATATTTATTTATGTTTAGTGTG- GTTTGAAATTTACCTCTTTTCAATGTTTTAA 1796
QY 2368 -----GTGCTCTGACATAAATATACATGACTATGTTTAAAGAGGCTTT 2408
DB 1797 ATAAAGTGAATTTATGATGTTTGTATACATGATGATTTATGTTTAAAGAGGCTTT 1856

```



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 20:46:55 ; Search time 11414 Seconds  
(without alignments)  
13188.614 Million cell updates/sec

Title: US-10-644-084-1  
Perfect score: 2692  
Sequence: 1 cgtagagagtcacagagc.....ttgattatggaqtaatqggg 2692

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 48236798 seqs, 27959665780 residues  
Total number of hits satisfying chosen parameters: 96473596

```

Minimum DB seq length: 0
Maximum DB seq length: 200000000

```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:★

```

1: gb_est1: *
2: gb_est3: *
3: gb_est4: *
4: gb_est5: *
5: gb_est6: *
6: gb_hic: *
7: gb_est2: *
8: gb_est7: *
9: gb_est8: *
10: gb_est9: *
11: gb_gse1: *
12: gb_gse2: *
13: gb_gse3: *
14: gb_gse4: *

```

SUMMARIES									
Result No.	Query #		Length	DB	ID	Description			
	Match	Score							
1	99.4	2676	3185	6	AK043865	Mus muscu			
2	99.4	2676	3413	6	AK031356	Mus muscu			
3	73.8	1985.4	2245	6	AK077112	Mus muscu			
4	73.3	1973	2371	6	AK049080	Mus muscu			
5	68.1	1832	1845	14	AY412493	Mus muscu			
6	63.2	1701.4	3154	6	BC035580	Homo sapi			
7	63.2	1701.4	63.2	5658	6	BSM801317	Homo sapi		
8	52.6	1416.6	52.6	2111	6	CR613292	full-length		
9	50.7	1363.8	50.7	1845	14	AY412491	Homo sapi		
10	50.4	1357.4	50.4	1845	14	AY412492	Homo sapi		
11	32.6	876.4	32.6	878	9	CX565842	Pan trogl		
12	30.2	812	30.2	840	5	CF5659831	UI-M-IB0-		
13	30.1	810	30.1	834	9	CF949831	UI-M-H30-		
14	29.8	803	29.8	804	8	CX562093	UI-M-FC0-		
15	29.6	798	29.6	798	9	CN539511	UI-M-HU0-		
16	29.3	789	29.3	822	8	CX562859	UI-M-IB0-		
17	28.5	768.4	28.5	771	3	CX563362	UI-M-FC0-		
18	28.3	761	28.3	933	4	BQ179212	UI-M-EW0-		
19	28.0	753.2	28.0	771	5	CB205832	AGNC002		
20	28.0	753.2	28.0	771	5	CF737129	UI-M-HD0-		

20	750.2	27.9	770	5	CF729394	UI-M-HD0-
21	749	27.8	760	8	CN528686	UI-M-HQ0-
22	748.2	27.8	845	3	BQ769157	UI-M-FC0-
23	739.2	27.5	746	8	CN528831	UI-M-HQ0-
24	721.4	26.8	723	5	CF737957	UI-M-HD0-
25	710.4	26.4	712	5	CF732776	UI-M-GZ0-
26	690	25.6	736	8	CX236608	MBM02133
27	679	25.2	679	8	CN532570	UI-M-HQ0-
28	677.8	25.2	683	8	CN424815	UI-M-HQ0-
29	675	25.1	687	9	CX568194	UI-M-IB0-
30	658	24.4	903	10	DV893190	LB02726.C
31	653.2	24.3	1005	1	AL582610	AL582610
32	643	23.9	643	3	BQ085173	H3109D09-
33	638	23.7	640	5	CF737822	UI-M-HD0-
34	633	23.5	833	8	CN117234	AGENCOURT
35	623.4	23.2	638	8	CN720771	E0809F12-
36	622.6	23.1	954	1	AL561392	AL561392
37	622.2	23.1	635	5	CF899025	A0247H06-
38	616.4	22.9	860	10	D7831879	LB00413.C
39	614.4	22.8	633	1	AV279314	AV279314
40	614.2	22.8	805	2	BQ342849	BQ342849
41	614.2	22.8	828	10	DV803752	LB035829
42	612.2	22.7	658	7	BB614736	BB614736
43	606.8	22.5	623	4	CA548100	C0801B11-
44	601.2	22.3	752	7	BF539334	602049514
45	600	22.3	600	2	BG804175	0252-45 M

## ALIGNMENTS

[illegible]

ACCESSION	AK043865
VERSION	AK043865.1
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL  
PUBMED  
10349636

2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, F., Muramatsu, M. and Hayashizaki, Y. Prepare full-length cDNA libraries for cap-trapper-selected cDNAs to normalize and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	
TITLE	
JOURNAL	
PUBLISHED	11042159

REFERENCE	AUTHORS	TITLE	JOURNAL	PUBMED
3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	RIKEN integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer	Genome Res. 10 (11), 1757-1771 (2000)	11076861

REFERENCE	AUTHORS	TITLE
4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	Functional annotation of a full-length mouse cDNA collection

JOURNAL REFERENCE AUTHORS	Nature 409, 685-690 (2001)
TITLE	The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002)
TITLE	RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium. Antisense Transcription in the Mammalian Transcriptome
JOURNAL REFERENCE AUTHORS	Science 309, 1559-1563 (2005)
TITLE	The PANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The Transcriptional Landscape of the Mammalian Genome
JOURNAL REFERENCE AUTHORS	Science 309, 1559-1563 (2005)
TITLE	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hirozane, T., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission
JOURNAL REFERENCE AUTHORS	Submitted (16-JUL-2001) Yoshihide Havaehizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
TITLE	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
FEATURES source	URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers 1. .3185 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:A830043F14" /db_xref="taxon:10090" /clone="A830043F14" /tissue_type="cortex" /clone_libs="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days neonate" 389. .2233 /note="unnamed protein product; HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus] (SPT AAH21749, evidence: FASTY, 99.8%ID, 100%length, match=1842) putative" /codon_start=1 /protein_id="BAC31684.1" /db_xref="GI:26335972" /translation="MGDMVTVDPLCTENKNLSQYTSKMSPPSYISQVLCSVP LSKNVGVGVCFTGIESISYLDQELTTFGPFSLYESKSEAKRELNIIVAVLNC MNEILLVKNLQAQESVETONLQSDMDHQAQKLEQLETSRRMIGLQDR QLOCKNSHOLLKNEVDQKLNIIASRATONHDVSKEREYVKLKERHQLVNV KDKNAMDVLNYGRADGKGSWRTEARNEDEMYKILLNDYERQKILMENAE LKKVLQMKMEISLLSPQKKPRAEADGTGTVASIDEDDSGELSRDSVWGUSCDT VREGUTNSIRKQWILKSHVEKLDNQASKVHSEGLNEEDVISRQDHQETKEULEIE RCKEMIKAQQQLLQQLATTCDDDTTSLRLDCYLLLEERLKEEWTLFKEQKNFERER
CDs	

	polyA_signal	polyA_site	ORIGIN
	3168..3173	3185	
	/note="putative"	/note="putative"	
	Query Match	Score 2676;	DB 6; Length 3185;
	Best Local Similarity	99.4%;	Pred. No. 0;
	Matches 2689;	Conservative	0; Mismatches
			0; Indels
			3; Gaps
			1;
Qy	1	CGTAGGAGCTGACAGGAGCTGTTGTAAGCGTGCAGCAGCTGAGCCGCCCTCCTCAGGTAT	60
Db	310	CGTAGGAGCTGACAGGAGCTGTTGTAAGCGTGCAGCAGCTGAGCCGCCCTCCTCAGGTAT	369
Qy	61	CTCGGCTCTCGAACTTTGCTATGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	120
Db	370	CTCGGCTCTCGAACTTTGCTATGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTAC	429
Qy	121	AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCGTCCAGTTTGTGA	180
Db	430	AGAAACAAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCGTCCAGTTTGTGA	489
Qy	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTTATCCAAAACGTCATGGTGTTCGCG	240
Db	490	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTTATCCAAAACGTCATGGTGTTCGCG	549
Qy	241	TGCTTTCTGCACAGGAGAGAACATTGAAACAAAGATTTCCTATCTTGCATCAGGAGCTGAC	300
Db	550	TGCTTTCTGCACAGGAGAGAACATTGAAACAAAGATTTCCTATCTTGCATCAGGAGCTGAC	609
Qy	301	CACCTTCGGTTTCCTTCCTTGTATGAGAAATCCAAAGATTAAGAGGCAAGAGAGAAAT	360
Db	610	CACCTTCGGTTTCCTTCCTTGTATGAGAAATCCAAAGATTAAGAGGCAAGAGAGAAAT	669
Qy	361	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTCTCGTGTCTCAGCGGAGAGAACCT	420
Db	670	AAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTCTCGTGTCTCAGCGGAGAGAACCT	729
Qy	421	GCTGGCCAGGAGAGCGTGGAGACACAACTTTGAAGCTGGGAGTGCATGAGACCACT	480
Db	730	GCTGGCCAGGAGAGCGTGGAGACACAACTTTGAAGCTGGGAGTGCATGAGACCACT	789
Qy	481	GCAGAGCTGTACGCCAACTTAAGGAGCTTGGAAACCTCCAGCGGGAGAGATGATCGG	540
Db	790	GCAGAGCTGTACGCCAACTTAAGGAGCTTGGAAACCTCCAGCGGGAGAGATGATCGG	849
Qy	541	GCTTCAAGAGAGAGACAGGAGCTGCAGTGCAGAAACAGAGATTTCATCAGCTCCTGAA	600
Db	850	GCTTCAAGAGAGAGACAGGAGCTGCAGTGCAGAAACAGAGATTTCATCAGCTCCTGAA	909
Qy	601	GAATGAGAAAGATGAGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA	660
Db	910	GAATGAGAAAGATGAGTACAAAATTAACAAAATATCATAGCCAGCGGGCTACTCAGTA	969
Qy	661	TAATCATGATGTAAGAGGAGGAGCGTGAATTAATAAGCTAAAGAGGCGCTGCATCA	720
Db	970	TAATCATGATGTAAGAGGAGGAGCGTGAATTAATAAGCTAAAGAGGCGCTGCATCA	1029
Qy	721	GCTCGTTTATGAAACAAAGAGATATAAATCATAGCCATGGATGTTTTTAAATTTATGTGGT	780
Db	1030	GCTCGTTTATGAAACAAAGAGATATAAATCATAGCCATGGATGTTTTTAAATTTATGTGGT	1089
Qy	781	AGCTGATGCAACAGAGGCTCATGAGGACTGACAAAACAGAACCCAGGAATGAAGATGA	840
Db	1090	AGCTGATGCAACAGAGGCTCATGAGGACTGACAAAACAGAACCCAGGAATGAAGATGA	1149
Qy	841	GATGTACAAAATCTCTGTTGAATGATTAAGTACCGCCAGAGAGCAGATCTCTGATGGAGAA	900
Db	1150	GATGTACAAAATCTCTGTTGAATGATTAAGTACCGCCAGAGAGCAGATCTCTGATGGAGAA	1209



QY 901 CGCGAGCTGAAGAAGTCTCCAGCAGATGAGAGGAGATGATCTCTCTCTCTCTCC 960  
Db 1210 CGCGAGCTGAAGAAGTCTCCAGCAGATGAGAGGAGATGATCTCTCTCTCTCTCC 1269  
QY 961 TCAGAGAAAGAGCCAGGGAAGAGCAGAGGACGGCAGAGCACTGTGCTATCTCCGA 1020  
Db 1270 TCAGAGAAAGAGCCAGGGAAGAGCAGAGGACGGCAGAGCACTGTGCTATCTCCGA 1329  
QY 1021 TATAGAGATGATCTCTGGGGAACGTAGCAGAGACAGCGTGTGGGCGCTTCTCTGACAC 1080  
Db 1330 TATAGAGATGATCTCTGGGGAACGTAGCAGAGACAGCGTGTGGGCGCTTCTCTGACAC 1389  
QY 1081 TGTGAGAGAGCAGCTGACAAAAGCAGATCAGGAAACAGTGGAGAAATTTGAAAAGTCATGT 1140  
Db 1390 TGTGAGAGAGCAGCTGACAAAAGCAGATCAGGAAACAGTGGAGAAATTTGAAAAGTCATGT 1449  
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGACGT 1200  
Db 1450 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGACGT 1509  
QY 1201 CATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGACGGTG 1260  
Db 1510 CATCTCAGCAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGACGGTG 1569  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGTGTGA 1320  
Db 1570 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTAAAGGCTCTTAAAGGCTCTTAAAGGCTCTTAA 1626  
QY 1321 TGATGACACCACTCACTGTTCGAGAGCTGTACTTGTCTGGAAGAAAAGGAAAGCGCTTAA 1380  
Db 1627 TGATGACACCACTCACTGTTCGAGAGCTGTACTTGTCTGGAAGAAAAGGAAAGCGCTTAA 1686  
QY 1381 AGAAGATGGACCTTTTAAAGAGCAAAAAGAAATTTGAGAGAGAAAGCGCAAGCTT 1440  
Db 1687 AGAAGATGGACCTTTTAAAGAGCAAAAAGAAATTTGAGAGAGAAAGCGCAAGCTT 1746  
QY 1441 TACGAAGCTGCCATTCGATGGGCTTGGAGAGAAAGCGTTTGAAGAGCGAGCCAG 1500  
Db 1747 TACGAAGCTGCCATTCGATGGGCTTGGAGAGAAAGCGTTTGAAGAGCGAGCCAG 1806  
QY 1501 CTGGTAAAGCAGCAGTGTAAAAATGACAGAACTTTGACACCAAGAACTCAGAAAATGT 1560  
Db 1807 CTGGTAAAGCAGCAGTGTAAAAATGACAGAACTTTGACACCAAGAACTCAGAAAATGT 1866  
QY 1561 GAACTTTTCACTGCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAG 1620  
Db 1867 GAACTTTTCACTGCTCTCAGGAAGTTCTGATCCAGACAATCTTATAGTCCACTCAG 1926  
QY 1621 GCCAGGCANAGAGCTACAGAGTGGCTAATGGGTGCCAGCTTGCACATCAAACT 1680  
Db 1927 GCCAGGCANAGAGCTACAGAGTGGCTAATGGGTGCCAGCTTGCACATCAAACT 1986  
QY 1681 GACTAAATCTCTCTGCTCCTCAGTCTTCTACTTCTGAGTCTGAGCAGACATTCATGTGT 1740  
Db 1987 GACTAAATCTCTCTGCTCCTCAGTCTTCTACTTCTGAGTCTGAGCAGACATTCATGTGT 2046  
QY 1741 GTCTGAACACAGTTCATCAGTGTGCTGAATATAAATCTCCTGAAGAAAGTAAACCAAGTGA 1800  
Db 2047 GTCTGAACACAGTTCATCAGTGTGCTGAATATAAATCTCCTGAAGAAAGTAAACCAAGTGA 2106  
QY 1801 GGTTCAGAGAGAAAGCAGGATCAGAGTGGAGCGTGCAGTGGAGCCAGCTCGCGGA 1860  
Db 2107 GGTTCAGAGAGAAAGCAGGATCAGAGTGGAGCGTGCAGTGGAGCCAGCTCGCGGA 2166  
QY 1861 GGGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCAGCGGGACCGAGATGACTT 1920  
Db 2167 GGGGTGCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCAGCGGGACCGAGATGACTT 2226  
QY 1921 ACCTTAAATGTGCGGGCTGCAAGTGTCTTCCAGATGTGCGCTAGAGAGATTGACACAGG 1980  
Db 2227 ACCTTAAATGTGCGGGCTGCAAGTGTCTTCCAGATGTGCGCTAGAGAGATTGACACAGG 2286  
QY 1981 GTGTAGCATAAAGTCAGTCGCTAACTTAAGATGCTCAGAGTTGTTGTTGGACTTCGC 2040

Db 2287 GTGTAGCATAAAGTCAGTCGCTAACTTAAGATGCTCAGAGTTGTTGTTGGACTTCGC 2346  
QY 2041 TGTCTTCCCCCAAGAGCTGAAATGCTTAAGCTTACTTAAAGGATGCAAAAGCTTTGGTTGT 2100  
Db 2347 TGTCTTCCCCCAAGAGCTGAAATGCTTAAGCTTACTTAAAGGATGCAAAAGCTTTGGTTGT 2406  
QY 2101 GTGTTAGTAACAGAAAGCCCTGGCTCTGTGACTGCGAGAAATGCAATGGGCTTTGGATGGA 2160  
Db 2407 GTGTTAGTAACAGAAAGCCCTGGCTCTGTGACTGCGAGAAATGCAATGGGCTTTGGATGGA 2466  
QY 2161 ACAGAGCGCTCGAATGATTCGCTCCAGAGTACCGAAGAGCACTTTTAGGGAAGTGT 2220  
Db 2467 ACAGAGCGCTCGAATGATTCGCTCCAGAGTACCGAAGAGCACTTTTAGGGAAGTGT 2526  
QY 2221 TCCTGTAACACATTAATATTCGTCCTCAAGTGTGGTTGGCATTTGGAAGTGTAGCTTTACT 2280  
Db 2527 TCCTGTAACACATTAATATTCGTCCTCAAGTGTGGTTGGCATTTGGAAGTGTAGCTTTACT 2586  
QY 2281 TGAATGATATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTATTTAGTGTGAT 2340  
Db 2587 TGAATGATATCTGTAGATTTTAAACAAAGCAGGTTCTATATTTATTTAGTGTGAT 2646  
QY 2341 TTTGGGATTACCTCTTTTCATATGTTTGTCTGTGTACATAAATACATGACTATGTTAA 2400  
Db 2647 TTTGGGATTACCTCTTTTCATATGTTTGTCTGTGTACATAAATACATGACTATGTTAA 2706  
QY 2401 GAGGCTTTAAGGTTTAAACCTTACACCATCTGTTGAGTATAGCATTTTCATGCCAATTAA 2460  
Db 2707 GAGGCTTTAAGGTTTAAACCTTACACCATCTGTTGAGTATAGCATTTTCATGCCAATTAA 2766  
QY 2461 AATGTTTTCAGTGGCATGGTGTTCACAGAGTTTAGGACCCTGCCACATGACAGTTAAGA 2520  
Db 2767 AATGTTTTCAGTGGCATGGTGTTCACAGAGTTTAGGACCCTGCCACATGACAGTTAAGA 2826  
QY 2521 CTTTATTTTAAAGCCTCGGCAATAAAAATTCAGAGCTTCAAGCCCTTCATAGCTGAGTTGAG 2580  
Db 2827 CTTTATTTTAAAGCCTCGGCAATAAAAATTCAGAGCTTCAAGCCCTTCATAGCTGAGTTGAG 2886  
QY 2581 ATAACTAGAACTACTTAACGTTTACATTTTGAAGATTTTAAAGCATTTGATTTTATTTAT 2640  
Db 2887 ATAACTAGAACTACTTAACGTTTACATTTTGAAGATTTTAAAGCATTTGATTTTATTTAT 2946  
QY 2641 ATATGTAATGTTTAAATTTCTAAGAGAAATTTGATTTAGGATTAATGGG 2692  
Db 2947 ATATGTAATGTTTAAATTTCTAAGAGAAATTTGATTTAGGATTAATGGG 2998

RESULT 2  
AK031356  
LOCUS  
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length enriched library, clone:603041112 product:hypothetical protein, full insert sequence.  
ACCESSION AK031356  
VERSION AK031356.1 GI:26327238  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)





**KEYWORDS**  
**SOURCE** HTC; CAP trapper.  
**ORGANISM** Mus musculus (house mouse)  
 Mus musculus  
 Mus musculus

**REFERENCE**  
**AUTHORS** Carninci, P. and Hayashizaki, Y.  
**TITLE** High-efficiency full-length cDNA cloning  
**JOURNAL** Meth. Enzymol. 303, 19-44 (1999)  
**PUBMED** 10349636  
**REFERENCE**

**AUTHORS** Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL** Genome Res. 10 (10), 1617-1630 (2000)  
**PUBMED** 11042159  
**REFERENCE**

**AUTHORS** Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
**TITLE** RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
**JOURNAL** Genome Res. 10 (11), 1757-1771 (2000)  
**PUBMED** 11076861  
**REFERENCE**

**AUTHORS** The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL** Nature 409, 685-690 (2001)  
**PUBMED** 11706861  
**REFERENCE**

**AUTHORS** The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team.  
**TITLE** Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
**JOURNAL** Nature 420, 563-573 (2002)  
**PUBMED** 12106861  
**REFERENCE**

**AUTHORS** RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium.  
**TITLE** Antisense Transcription in the Mammalian Transcriptome  
**JOURNAL** Science 309, 1564-1566 (2005)  
**PUBMED** 16106861  
**REFERENCE**

**AUTHORS** The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).  
**TITLE** The Transcriptional Landscape of the Mammalian Genome  
**JOURNAL** Science 309, 1559-1563 (2005)  
**PUBMED** 16106861  
**REFERENCE**

**AUTHORS** Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirose, T., Hirozane, T., Hori, F., Imotani, K., Itoh, M., Itoh, K., Kaga, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
**REFERENCE**

**COMMENT** cDNA library was prepared and sequenced in Mouse Genome

---

**FEATURES**  
**source**

Location/Qualifiers  
 1..2245  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:49333406N16"  
 /db\_xref="taxon:10090"  
 /clone="49333406N16"  
 /sex="male"  
 /tissue\_type="testis"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 1..2245  
 /note="HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus] (SPTR|AAH21749, evidence: FASTY, 99.7%ID, 100%length, match=1842)  
 putative"

**misc\_feature**

1..2245  
 /note="HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus] (SPTR|AAH21749, evidence: FASTY, 99.7%ID, 100%length, match=1842)  
 putative"

**ORIGIN**

Query Match 73.8%; Score 1985.4; DB 6; Length 2245;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1999; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 1 CGTAGAGAGTGCAGGAGCTGTTGTAAGCGTGCAGCACTGAGCGCCCTCTCAGGTAT 60  
 DB 246 CGTAGAGAGTGCAGGAGCTGTTGTAAGCGTGCAGCACTGAGCGCCCTCTCAGGTAT 305  
 QY 61 CTGGGCTCTGAACTTCTGCTATGGAGATTGGATGACTGTGACAGATCCAGTCTCTGTATC 120  
 DB 306 CTGGGCTCTGAACTTCTGCTATGGAGATTGGATGACTGTGACAGATCCAGTCTCTGTATC 365  
 QY 121 AGAACAACAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGTCAGTTTGTGA 180  
 DB 366 AGAACAACAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGTCAGTTTGTGA 425  
 QY 181 CTCCAGCAAGTCTCTGCTCTTTCAGTACCTTTATCCAAAACGTCATGCTGTTTCGG 240  
 DB 426 CTCCAGCAAGTCTCTGCTCTTTCAGTACCTTTATCCAAAACGTCATGCTGTTTCGG 485  
 QY 241 TGTCTTCTGCACAGGAGAGAACATTGAAACAAAGATTTCTCTATCTTGATCAGGAGCTGAC 300  
 DB 486 TGTCTTCTGCACAGGAGAGAACATTGAAACAAAGATTTCTCTATCTTGATCAGGAGCTGAC 545  
 QY 301 CACCTTCGGGTTTCTCTTCTGTTGATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT 360  
 DB 546 CACCTTCGGGTTTCTCTTCTGTTGATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT 605  
 QY 361 AAATATAGTCTGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTCTCAGCGGAAGAACCT 420  
 DB 606 AAATATAGTCTGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTCTCAGCGGAAGAACCT 665  
 QY 421 GCTGGCCCGAGGAGAGCGTGAGACACAGAACTTTGAAGCTGGGAGTGCATGAGCCACCT 480  
 DB 666 GCTGGCCCGAGGAGAGCGTGAGACACAGAACTTTGAAGCTGGGAGTGCATGAGCCACCT 725  
 QY 481 GCAGAGCTGTACGCCAACTTAAGGAGCAGTTGGAAACCTCCAGGGCGGAGATGATCGG 540  
 DB 726 GCAGAGCTGTACGCCAACTTAAGGAGCAGTTGGAAACCTCCAGGGCGGAGATGATCGG 785  
 QY 541 GCTTCAAGAGAGAGACAGGAGCTGCAGTGCAAGAACAGAGATTTGCATCAGTCTCTGAA 600  
 DB 786 GCTTCAAGAGAGAGACAGGAGCTGCAGTGCAAGAACAGAGATTTGCATCAGTCTCTGAA 845  
 QY 601 GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGCTACTCAGTA 660  
 DB 846 GAATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAGCGGCTACTCAGTA 905

QY 661 TAATCATGTGTAAGAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 720  
DB |||||  
QY 906 TAATCATGTGTAAGAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCA 965  
DB |||||  
QY 721 GCTCGTTATGAACAAGAGTAAATAATAGCCATGGATGTTTAAATATATGCGGTCG 780  
DB |||||  
QY 966 GCTCGTTATGAACAAGAGTAAATAATAGCCATGGATGTTTAAATATATGCGGTCG 1025  
DB |||||  
QY 781 AGCTGATGCAAAACGAGGCTCATGAGGACTGACAAAACAGAACCCAGGAATGAAGATGA 840  
DB |||||  
QY 1026 AGCTGATGCAAAACGAGGCTCATGAGGACTGACAAAACAGAACCCAGGAATGAAGATGA 1085  
DB |||||  
QY 841 GATGTACAAAATCTGTTCAATGATATGAGTACCGCCAGAGCAGATCCTGATGAGAA 900  
DB |||||  
QY 1086 GATGTACAAAATCTGTTCAATGATATGAGTACCGCCAGAGCAGATCCTGATGAGAA 1145  
DB |||||  
QY 901 CGCGAGGCTGAAGAAAGGCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCC 960  
DB |||||  
QY 1146 CGCGAGGCTGAAGAAAGGCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCC 1205  
DB |||||  
QY 961 TCAGAAGAGAAAGCCAGGAAAGAGCAGAGGACGGCAAGGCACTGTTGCTATCTCCGA 1020  
DB |||||  
QY 1206 TCAGAAGAGAAAGCCAGGAAAGAGCAGAGGACGGCAAGGCACTGTTGCTATCTCCGA 1265  
DB |||||  
QY 1021 TATAGAAGTACTCTGGGAACTGAGCAGAGCAGACGCTGTGGGCGCTTCTCTGTGACAC 1080  
DB |||||  
QY 1266 TATAGAAGTACTCTGGGAACTGAGCAGAGCAGACGCTGTGGGCGCTTCTCTGTGACAC 1325  
DB |||||  
QY 1081 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGATTTTGAAGAAATCATGT 1140  
DB |||||  
QY 1326 TGTGAGAGAGAGCTGACAAAACAGCATCAGGAAACAGTGGAGATTTTGAAGAAATCATGT 1385  
DB |||||  
QY 1141 AGAAAACTCGATAACCAAGCTTCCGAGGTACACTCAGAGGCGCTTAATAGGAGAGCGT 1200  
DB |||||  
QY 1386 AGAAAACTCGATAACCAAGCTTCCGAGGTACACTCAGAGGCGCTTAATAGGAGAGCGT 1445  
DB |||||  
QY 1201 CATCTCAGCAGACCAATGAGCAGACACTGAGAACTGAGCTGGAGATTTAGCGGTG 1260  
DB |||||  
QY 1446 CATCTCAGCAGACCAATGAGCAGACACTGAGAACTGAGAGCTGGAGATTTAGCGGTG 1505  
DB |||||  
QY 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTA 1320  
DB |||||  
QY 1506 TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTA 1562  
DB |||||  
QY 1321 TGATGACACCACTCAGCTGTTCCGAGACTGTTACTTGTCTGGAAGAAAAGGAAACGCTTAA 1380  
DB |||||  
QY 1563 TGATGACACCACTCAGCTGTTCCGAGACTGTTACTTGTCTGGAAGAAAAGGAAACGCTTAA 1622  
DB |||||  
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1440  
DB |||||  
QY 1623 AGAAGAGTGGACCTTTTAAAGAGCAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT 1682  
DB |||||  
QY 1441 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG 1500  
DB |||||  
QY 1683 TACAGAAGCTGCCATTCGATTTGGGTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG 1742  
DB |||||  
QY 1501 CTGGGTAAAGCAGCAGTGTAAAAATGACGAACCTTTGACCAACCAAGAACTCAGAAAAATGT 1560  
DB |||||  
QY 1743 CTGGGTAAAGCAGCAGTGTAAAAATGACGAACCTTTGACCAACCAAGAACTCAGAAAAATGT 1802  
DB |||||  
QY 1561 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCAGC 1620  
DB |||||  
QY 1803 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCGATCCAGACAATCTTATAGTCCACTCAGC 1862  
DB |||||  
QY 1621 GCCAGGCAAAAGAGCTTACACAGTGTGGCTTAATGGGTGCGAGCTTGCACATCAAACT 1680  
DB |||||  
QY 1863 GCCAGGCAAAAGAGCTTACACAGTGTGGCTTAATGGGTGCGAGCTTGCACATCAAACT 1922  
DB |||||  
QY 1681 GACTAAATCTCTTCTGCTTCCACCTTCTACTTTCAGACTTTCCGAGAGACATTTATGTGT 1740  
DB |||||  
QY 1923 GACTAAATCTCTTCTGCTTCCACCTTCTACTTTCAGACTTTCCGAGAGACATTTATGTGT 1982  
DB |||||

QY 1741 GTCTGAACACAGTTCATCTCAGTGTGCTGAATATATACTCTCTGAAGAAAGTAAACCAAGTGA 1800  
DB |||||  
QY 1801 GGTTCGAAGAGAAAGCAGGATCAGAACTGGAGCGTGCAGTCGAGGCCCGAGCTCCGGGA 1860  
DB |||||  
QY 2043 GGTTCGAAGAGAAAGCAGGATCAGAACTGGAGCGTGCAGTCGAGGCCCGAGCTCCGGGA 2102  
DB |||||  
QY 1861 GGGGTGCTACAGCGGATGCTCTCTCGGCTTCCAGGAGCGCTCACGGGGACCCGAGATGACTT 1920  
DB |||||  
QY 2103 GGGGTGCTACAGCGGATGCTCTCTCGGCTTCCAGGAGCGCTCACGGGGACCCGAGATGACTT 2162  
DB |||||  
QY 1921 ACCTTAAATGTGCGGCTGCGAGTGTCTTCCAGATGTGCGCTAGAGAGTTGACACAGG 1980  
DB |||||  
QY 2163 ACCTTAAATGTGCGGCTGCGAGTGTCTTCCAGATGTGCGCTAGAGAGTTGACACAGG 2222  
DB |||||  
QY 1981 GTGTAGCATATAAGTCAGTCGTCT 2003  
DB |||||  
QY 2223 GTGTAGCATATAAGTCAGTCGTCT 2245  
DB |||||  
RESULT 4  
AK049080  
LOCUS  
DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone: C230097E14 product: hypothetical protein,  
full insert sequence.  
ACCESSION AK049080  
VERSION AK049080.1 GI:26339745  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
PUBMED 10349636  
REFERENCE  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, N., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED 11042159  
REFERENCE  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichannel sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED 11076861  
REFERENCE  
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
PUBMED 11076861  
REFERENCE  
AUTHORS The FANTOM Consortium, the RIKEN Genome Exploration Research Group  
Phase I and II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
PUBMED 11076861  
REFERENCE  
AUTHORS RIKEN Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium.

TITLE	Antisense Transcription in the Mammalian Transcriptome	Db	358	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTGCAGCACTGAGCCGCTCCTCAGGTAT	417
JOURNAL	Science 309, 1564-1566 (2005)	Qy	61	CTTGCTCTGGAACCTGCTATGGAGATTCGATGACAGATCAGTTCCTGTGTAC	120
REFERENCE	The FANTOM Consortium, Riken Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The Transcriptional Landscape of the Mammalian Genome	Db	418	CCTGGCTCTGGAACCTGCTATGGAGATTCGATGACAGATCAGTTCCTGTGTAC	477
JOURNAL	Science 309, 1559-1563 (2005)	Qy	121	AGAAACAAAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGCTCCAGTTGTA	180
REFERENCE	8 (bases 1 to 2371)	Db	478	AGAAACAAAAATCTCTCAATATACCTCAGAAACAAAGATGCTCCGCTCCAGTTGTA	537
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.	Qy	181	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTATCCAAACAAAGTGCATGGTTCGG	240
TITLE	Direct Submission	Db	538	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTATCCAAACAAAGTGCATGGTTCGG	597
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)	Qy	241	TGCTCTTCGACAGGAGAGCAATGAAACAAAGTATTTCTTATCTTGCATCAGGAGCTGAC	300
REFERENCE	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.	Db	598	TGCTCTTCGACAGGAGAGCAATGAAACAAAGTATTTCTTATCTTGCATCAGGAGCTGAC	657
AUTHORS	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Qy	301	CACCTTCGGGTTTCCTTCTTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT	360
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	658	CACCTTCGGGTTTCCTTCTTGTATGAAGAAATCCAAAGTAAAGAGGCAAGAGAGAAAT	717
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	361	AAATATAGTCTGCTGTTCTGAACTGTATGAACGAGCTCTCTGCTTCAGCGGAAGAACCT	420
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Db	718	AAATATAGTCTGCTGTTCTGAACTGTATGAACGAGCTCTCTGCTTCAGCGGAAGAACCT	777
AUTHORS	prepare mouse tissues. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Qy	421	GCTGGCCCGAGAGAGCGTGAGACACAGAACTTGAAGCTGGGCAAGTACATGGACCACT	480
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	778	GCTGGCCCGAGAGAGCGTGAGACACAGAACTTGAAGCTGGGCAAGTACATGGACCACT	837
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	481	GCAGAGCTGCTAGCCCAAACTTAAAGGAGCAGTTGGAAACCTCCAGCGGGAGATGATCGG	540
AUTHORS	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Db	838	GCAGAGCTGCTAGCCCAAACTTAAAGGAGCAGTTGGAAACCTCCAGCGGGAGATGATCGG	897
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	541	GCTTCAAGAGAGACAGGAGCTGCGAGTCAAGAAACAGAGAGTTTGCATCAGCTCTCGAA	600
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	898	GCTTCAAGAGAGACAGGAGCTGCGAGTCAAGAAACAGAGAGTTTGCATCAGCTCTCGAA	957
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	601	GAATGAGAAAGATGAGTACAAAAATTTACAAATATCATAGCCAGCGGGCTACTCAGTA	660
AUTHORS	prepare mouse tissues. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Db	958	GAATGAGAAAGATGAGTACAAAAATTTACAAATATCATAGCCAGCGGGCTACTCAGTA	1017
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	661	TAATCATGATGTGAAGAGAGGAGCGTGAATATAATGAAGCTAAAGGAGCGCTGCATCA	720
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	1018	TAATCATGATGTGAAGAGAGGAGCGTGAATATAATGAAGCTAAAGGAGCGCTGCATCA	1077
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	721	GCTCGTTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGTCTG	780
AUTHORS	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Db	1078	GCTCGTTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTTAAATTTATGTGGTCTG	1137
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	781	AGCTGATGGCAACAGGCGCTCATGGAGGACTGCACAAACAGAACCCAGGAAATGAAGATGA	840
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	1138	AGCTGATGGCAACAGGCGCTCATGGAGGACTGCACAAACAGAACCCAGGAAATGAAGATGA	1197
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	841	GATGTACAAAATCTGTTGTAATGATTAAGTACCGCCAGAGAGCAGATCCTGTATGGAGAA	900
AUTHORS	prepare mouse tissues. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Db	1198	GATGTACAAAATCTGTTGTAATGATTAAGTACCGCCAGAGAGCAGATCCTGTATGGAGAA	1257
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	901	CGCGAGCTGAAGAGAGTCTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCC	960
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	1258	CGCGAGCTGAAGAGAGTCTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCC	1317
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	961	TCAGAAAG	1020
AUTHORS	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers	Db	1318	TCAGAAAG	1377
TITLE	Encyclopedia Project of Genome Exploration Research Group in Riken	Qy	1021	TATGAGAGAGTACTCTCGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTTCTGTGACAC	1080
JOURNAL	Encyclopedia Project of Genome Exploration Research Group in Riken	Db	1378	TATGAGAGAGTACTCTCGGGAACCTGAGCAGAGACAGCGTGTGGGCGCTTTCTGTGACAC	1437
REFERENCE	Genomic Sciences Center and Genome Science Laboratory in RIKEN	Qy	1081	TGTGAGAGAGAGTGAACAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGTCAATGT	1140

ORIGIN

Query Match 73.3%; Score 1973; DB 6; Length 2371;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 2005; Conservative 0; Mismatches 0; Indels 12; Gaps 2;  
 Qy 1 CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTGCAGCACTGAGCCGCTCCTCAGGTAT 60



Db	1438	TGTGAGAGCAGCTGACAAACAGCATCAGGAAACAGTCGGAGATTTTGAAGAATCATGT	1497
Qy	1141	AGAAAACTCGATAACCAAGCTTCGAAAGTACACTCAGAGGCCCTTAATGAGGAGACGT	1200
Db	1498	AGAAAACTCGATAACCAAGCTTCGAAAGTACACTCAGAGGCCCTTAATGAGGAGACGT	1557
Qy	1201	CATCTCACGACAAAGACCATGACAGAGACTCAGAAACTCGAGCTGGAGATTGACGGTG	1260
Db	1558	CATCTCACGACAAAGACCATGACAGAGACTGAGAACTGGAGCTGGAGATTGACGGTG	1617
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACCGTGTGA	1320
Db	1618	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTAA--CAGCAGCTGGCCACCACCGTGTGA	1674
Qy	1321	TGATCAGCACCACCTCACTGTTCCGAGACTGTTACTGTCTGGAAGAAAAGGAACGCCTTAA	1380
Db	1675	TGATCAGCACCACCTCACTGTTCCGAGACTGTTACTGTCTGGAAGAAAAGGAACGCCTTAA	1734
Qy	1381	AGAAGAGTGGACCCCTTTTAAAGACCAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT	1440
Db	1735	AGAGAGTGGACCCCTTTTAAAGACCAAAAAGAAATTTTGAGAGAGAAAGCGGAAGCTT	1794
Qy	1441	TACAGAAGCTGTCATTCGATTGGGCTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG	1500
Db	1795	TACAGAAGCTGTCATTCGATTGGGCTTGGAGAGAAAGCGTTTGAAGAGAGCGAGCCAG	1854
Qy	1501	CTGGGTAAGCAGCAGTGTAAAAATGACGCACTTTGACCAACCAAGAACTCAGAAAAATGT	1560
Db	1855	CTGGGTAAGCAGCAGTGTAAAAATGACGCACTTTGACCAACCAAGAACTCAGAAAAATGT	1914
Qy	1561	GAACCTTTTCAGTGCCCTCTCAGGAA-----GTTCTGATCCAGCAATCTTATAGT	1611
Db	1915	GAACCTTTTCAGTGCCCTCTCAGGAAAGTAAACCAAGGTTCTGATCCAGCAATCTTATAGT	1974
Qy	1612	CCACTCAGCGGCCACGGCAAAAGAAAGTACACAGTGTGGCTAATGGGGTGCAGCTTTGCAC	1671
Db	1975	CCACTCAGCGGCCACGGCAAAAGAAAGTACACAGTGTGGCTAATGGGGTGCAGCTTTGCAC	2034
Qy	1672	ATCAAACTGACTAAATCTCTTCGCTCCTACCTTCTACTTCAGACTTTCCGCGAGACACA	1731
Db	2035	ATCAAACTGACTAAATCTCTTCGCTCCTACCTTCTACTTCAGACTTTTCGCGAGACACA	2094
Qy	1732	TTCATGTGTCTGAACACAGTTCATCAGTGTGCTGTAATTAATCTCTGGAAGAACTAA	1791
Db	2095	TTCATGTGTCTGAACACAGTTCATCAGTGTGCTGTAATTAATCTCTGGAAGAACTAA	2154
Qy	1792	ACCAAGTGAGGTTGCAAGAGAAAGCACCGATCAGAAAGTGGAGCGTGCAGTCGAGGCCAG	1851
Db	2155	ACCAAGTGAGGTTGCAAGAGAAAGCACCGATCAGAAAGTGGAGCGTGCAGTCGAGGCCAG	2214
Qy	1852	CTCGGGAGGGGTGCTACAGCGGATGCTCTCGGCCCTTCAGGAGCGCTCACGGGACCG	1911
Db	2215	CTCGGGAGGGGTGCTACAGCGGATGCTCTCGGCCCTTCAGGAGCGCTCACGGGACCG	2274
Qy	1912	AGATGACTTACCTTAAATGCGGGCTGCAGTGTCTGTTCCAGATGTCGCTAGAGAGT	1971
Db	2275	AGATGACTTACCTTAAATGCGGGCTGCAGTGTCTGTTCCAGATGTCGCTAGAGAGT	2334
Qy	1972	TGACACAGGGGTGAGCATAAAAGTCAGTCTGCTTAACTT 2008	
Db	2335	TGACACAGGGGTGAGCATAAAAGTCAGTCTGCTTAACTT 2371	

RESULT 5  
AY412493  
LOCUS AY412493 1845 bp DNA linear GSS 12-DEC-2003  
DEFINITION Mus musculus HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY412493  
VERSION AY412493.1 GI:39769458  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE AUTHORS	1 (bases 1 to 1845) Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J., Adams, M.D. and Cargill, M. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios Science 302 (5652), 1960-1963 (2003) 14671302 2 (bases 1 to 1845) Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Perriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J., Adams, M.D. and Cargill, M. Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA This sequence was made by sequencing genomic exons and ordering them based on alignment.
TITLE	Location/Qualifiers
JOURNAL	1..1845
PUBMED	/organism="Mus musculus"
REFERENCE	/mol_type="genomic DNA"
AUTHORS	/db_xref="taxon:10090"
	<1..>1845
	/locus_tag="HCM4575"
TITLE	
JOURNAL	
COMMENT	
FEATURES	
source	
gene	
ORIGIN	
Query Match	68.1%; Score 1832; DB 14; Length 1845;
Best Local Similarity	99.8%; Pred. No. 0;
Matches 1845; Conservative	0; Mismatches 0; Indels 3; Gaps 1;
Qy	80 ATGGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAAAATCTCTCT 139
Db	1 ATGGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAAAATCTCTCT 60
Qy	140 CAATATACCTCAGAAAAAAGATGTCCTCGTCCAGTTGTACTCCACAGAAATCTGTGC 199
Db	61 CAATATACCTCAGAAAAAAGATGTCCTCGTCCAGTTGTACTCCACAGAAATCTGTGC 120
Qy	200 TCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGGTGTCTTCGCACAGGAGAG 259
Db	121 TCTTCAGTACCTTTATCCAAAAACGTGATGGTGTTCGGTGTCTTCGCACAGGAGAG 180
Qy	260 AACATTGAACAAAGTATTTCTATCTTTGATCAGAGCTGACCACTTCGGGTTCCTTCC 319
Db	181 AACATTGAACAAAGTATTTCTATCTTTGATCAGAGCTGACCACTTCGGGTTCCTTCC 240
Qy	320 TTGTATGAAGAATCCAAAAAGTAAAGAGGCAAGAGAGATTTAAATATATAGTCGTTCTG 379
Db	241 TTGTATGAAGATTCAAAAAGTAAAGAGGCAAGAGAGATTTAAATATATAGTCGTTCTG 300
Qy	380 AACTGTATGAACAGAGCTGCTCGTCTTCAGCGGAAGAACCTGTGCGCCAGGAGAGCGTG 439
Db	301 AACTGTATGAACAGAGCTGCTCGTCTTCAGCGGAAGAACCTGTGCGCCAGGAGAGCGTG 360
Qy	440 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGACACCACTGCGAGAGCTGTACGCCAA 499
Db	361 GAGACACAGAACTTGAAGCTGGGCGAGTGACATGACACCACTGCGAGAGCTGTACGCCAA 420
Qy	500 CTTAAGGAGCAGTTGGAAAAAGTCCAGCGGAGATGATCGGCTTCAAGAGAGAGACAGG 559
Db	421 CTTAAGGAGCAGTTGGAAAAAGTCCAGCGGAGATGATCGGCTTCAAGAGAGAGACAGG 480
Qy	560 CAGCTCAGTGCAGAAACAGGAGTTTGCATCAGCTCCTCTGAAGAAATGAGAAAGATGAGGTA 619
Db	481 CAGCTCAGTGCAGAAACAGGAGTTTGCATCAGCTCCTCTGAAGAAATGAGAAAGATGAGGTA 540
Qy	620 CAAAAATTACAAAATATCATAGCCAGCGGGCTACTCAGTATATATCATATGATGTAAGAGG 679

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

Db 541 CAAATAATACAAATATCATAGCCAGCGGCTACTCAGTATATATCATGTGTGAAGAGG 600  
Qy 680 AAGAGCGTGAATATATAGCTTAAGAGCGGCTGCATCAGCTCGTTATGAACAAGAG 739  
Db 601 AAGAGCGTGAATATATAGCTTAAGAGCGGCTGCATCAGCTCGTTATGAACAAGAG 660  
Qy 740 GATAAACAATAGCATGATGTTTAAATATATGTGGTGCAGCTGATGCAACAGAGGC 799  
Db 661 GATAAACAATAGCATGATGTTTAAATATATGTGGTGCAGCTGATGCAACAGAGGC 720  
Qy 800 TCATGGAGGACTGACAAAACAGAACCCAGGAATGAAGATGAGATGTACAAATTTCTGTTG 859  
Db 721 TCATGGAGGACTGACAAAACAGAACCCAGGAATGAAGATGAGATGTACAAATTTCTGTTG 780  
Qy 860 AATGATTATGATACCGCCAGAACAGATCTGTATGGAGAACCGGAGCTGAAGAAGTC 919  
Db 781 AATGATTATGATACCGCCAGAACAGATCTGTATGGAGAACCGGAGCTGAAGAAGTC 840  
Qy 920 CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCTCTCAGAAGAAGACCCAGG 979  
Db 841 CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGCTCTCTCAGAAGAAGACCCAGG 900  
Qy 980 GAAAGAGCAGAGGACGGCAGACGACCTGTTGCTATCTCCGATATAGAAGATGACTCTGGG 1039  
Db 901 GAAAGAGCAGAGGACGGCAGACGACCTGTTGCTATCTCCGATATAGAAGATGACTCTGGG 957  
Qy 1040 GAACTGAGCAGACAGGCTGTGGGCTTTCCTGTGTGACACTGTGAGAGAGCCTGACA 1099  
Db 958 GAACTGAGCAGACAGGCTGTGGGCTTTCCTGTGTGACACTGTGAGAGAGCCTGACA 1017  
Qy 1100 AACAGCATCAGGAACAGTGGAGAAATTTGAAAGAGTCACTGAGAAAACTCGATAACCAA 1159  
Db 1018 AACAGCATCAGGAACAGTGGAGAAATTTGAAAGAGTCACTGAGAAAACTCGATAACCAA 1077  
Qy 1160 GCTTCGAAGGTACACTCAGAGGGCTTAATCAGGAGGACGTCATCTCAGCAACAGACCAT 1219  
Db 1078 GCTTCGAAGGTACACTCAGAGGGCTTAATCAGGAGGACGTCATCTCAGCAACAGACCAT 1137  
Qy 1220 GAGCAAGAGACTGAGAAACTGAGCTGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCT 1279  
Db 1138 GAGCAAGAGACTGAGAAACTGAGCTGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCT 1197  
Qy 1280 CAGCAGAGCTCTTACAGCAGCAGCTGSCCACCACGCTGTGATGATGACACCACTCACTG 1339  
Db 1198 CAGCAGAGCTCTTACAGCAGCAGCTGSCCACCACGCTGTGATGATGACACCACTCACTG 1257  
Qy 1340 TTGCGAGACTGTTACTTGTGGAAGAAAGAACCGCTTTAAAGAGAGTGACCCCTTTT 1399  
Db 1258 TTGCGAGACTGTTACTTGTGGAAGAAAGAACCGCTTTAAAGAGAGTGACCCCTTTT 1317  
Qy 1400 AAAGAGCAAAAAAGAAATTTGAGAGAGAAAGGCGAAGCTTTACAGAAGCTGCCATTGCA 1459  
Db 1318 AAAGAGCAAAAAAGAAATTTGAGAGAGAAAGGCGAAGCTTTACAGAAGCTGCCATTGCA 1377  
Qy 1460 TTGGGGTTGGAGAGAAAGGCGTTTGAAGACAGCGACGCTGGTAAAGCAGCAGTTT 1519  
Db 1378 TTGGGGTTGGAGAGAAAGGCGTTTGAAGACAGCGACGCTGGTAAAGCAGCAGTTT 1437  
Qy 1520 TTAACATGACGAACCTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTCAGTGCCTTC 1579  
Db 1438 TTAACATGACGAACCTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTCAGTGCCTTC 1497  
Qy 1580 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCACTCAGCGCCACGGCAAAAGAGCTA 1639  
Db 1498 TCAGGAAGTTCTGATCCAGACAATCTTATAGTCACTCAGCGCCACGGCAAAAGAGCTA 1557  
Qy 1640 CACAGTGGGCTTAATGGGGTGCCAGCTGACATCAAAAACCTGACTAAATCTCTTCTCCGCC 1699  
Db 1558 CACAGTGGGCTTAATGGGGTGCCAGCTGACATCAAAAACCTGACTAAATCTCTTCTCCGCC 1617  
Qy 1700 TCACCTTCTACTTCAGACTTTTCGCAGACACATTCATGTGTGTGACACAGTTCCATC 1759  
Db 1618 TCACCTTCTACTTCAGACTTTTCGCAGACACATTCATGTGTGTGACACAGTTCCATC 1677

Qy 1760 AGTGTCTGAATATACTCTCTGAAGAAAGTAAACCAAGTGAGGTTGCAAGAGAAAGCACG 1819  
Db 1678 AGTGTCTGAATATACTCTCTGAAGAAAGTAAACCAAGTGAGGTTGCAAGAGAAAGCACG 1737  
Qy 1820 GATCAGAAGTGGAGCGTGCAGTGCAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 1879  
Db 1738 GATCAGAAGTGGAGCGTGCAGTGCAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 1797  
Qy 1880 TCCTCGGCTTTCAGGAGCGCTCAGGGGACCGAGATGACTTACCTTAA 1927  
Db 1798 TCCTCGGCTTTCAGGAGCGCTCAGGGGACCGAGATGACTTACCTTAA 1845

RESULT 6  
BC035580  
LOCUS  
DEFINITION  
mRNA (cDNA clone IMAGE:5575710) .  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
BC035580 3154 bp mRNA linear HTC 12-DEC-2005  
Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,  
mRNA (cDNA clone IMAGE:5575710) .  
BC035580  
BC035580.1 GI:22028261  
HTC.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homniidae; Homo.  
1 (bases 1 to 3154)  
Klausner,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F.,  
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Scapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Loquellano,N.A., Peters,G.J.,  
Carninci,P., Prange,C., Raha,S.S., Bosak,S.A., McEwan,P.J.,  
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Vallalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,  
Schnerich,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Mammalian Gene Collection Program team  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 3154)

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
CONSTRM  
JOURNAL  
REMARK  
COMMENT  
NIH MGC Project  
Direct Submission  
Submitted (31-JUL-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Bethesda, MD 20892-2590, USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Akhter N., Ayele, K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
Blakesley,R.W., Bouffard,G.G., Green,K., Brinkley,C., Brookes,S.,  
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Hashighi,P.,  
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,F., Legaspi,R.,  
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,



McDowell, J., Pearson, R., Stantripop, S., Thomas, P. J., Touchman, J. W.,  
Tsurgeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 69 Row: n Column: 9  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 51317360  
 This clone has the following problem: no 5' EST match.

FEATURES  
source

```
location/Qualifiers
1. .3154
  /organism="Homo sapiens"
  /mol_type="rRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5575710"
  /tissue_type="testis, embryo"
  /clone_lib="NIH_MGC_92"
  /lab_host="DH10B"
  /notes="Vector: pCMV-SPORT6"
```

## ORIGIN

Query Match 63.2%; Score 1701.4; DB 6; Length 3154;  
Best Local Similarity 80.7%; Pred. No. 0;  
Matches 2214; Conservative 0; Mismatches 471; Indels 60; Gaps 17;

Qy	1	CGTAGGAGAGTGACAGAGAGCTGTTGTTAAAGCGTGCACGACATGAGCCGCCCTCCTCAGGTAT	60
Db	7	CATAGAGAGTGACAGCAGCTAGACTAAATGTTTAACTGCTGAACTAGTTTCCCTCAGGTAT	66
Qy	61	CTTGGCTCTCGAACTTGTCTATGGGAGATTGGATGACTGTGCACAGATCCAGTTCGTGTGTCAC	120
Db	67	CTTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTACAGATCCAGGCTGTGTTCTTC	126
Qy	121	AGAAAACAAAAAATCTCTCTCAATATACCTCAGAAAACAAAGATGTCCTCGTCCAGTTTGTGA	180
Db	127	AGAAAGCAAACTATCTCTCAATATACCTCAGAAAACAAAGATGTCCTCAATCAAGTTTATA	186
Qy	181	CTCCAGCAAGTTCTGTGCTCTTCAGTACCTTTATCCAAAAACGTGCATGGTGTGTTTCGG	240
Db	187	CTCACAGCAAGTGTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAG	246
Qy	241	TGTCCTTCTGCACAGGAGAGAACATTGAACAAAGTATTTCTATCTTGTATCAGAGCTGAC	300
Db	247	TGCTCTTCTGCACAGAGATAATATTGAACAGAGTATCTCATATCTTGTATCAGGAATTGAC	306
Qy	301	CACCTTCGGGTTTCTCTTCTTGTATGAAGAATCCAAAGATTAAAGAGCAAGAGAGAAATT	360
Db	307	TACTTTTGGTTTTCCTTTCATTATATGAAGAAATCCAAAGGTAAGAGACAAGAGAGAGATT	366
Qy	361	AAATATAGTCGCTGTTCTCGAACTGTATGAACGAGCTGCTCGTGTCTCAGCGGGAAGAACCT	420
Db	367	AAATATAGTAGCTGTACTAAATTTGTATGATGAGCTGCTGTGTCTTACGCGGAAGAACCT	426
Qy	421	GCTGGCCGAGGAGCGTGGAGACACAGAACTTTGAAGCTGGGCGAGTGACATGACCACCT	480
Db	427	TCTAGCTCAGGAAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGTGATATGACCACCT	486
Qy	481	GCAGAGCTGCTACGCCAAACTTTAAGGAGCAGTTTGGAAAAGTCCAGCGGGGAGATGATCGG	540
Db	487	ACAGAGCTGCTACTCAAAAACCTTAAAGAACAACTTGGAAAAGTCCAGGAGGGAAATGATGG	546
Qy	541	GCCTTCAAGAGAGACAGGCGAGCTGCAAGTGCACAGAAACAGGAGTTTGCATCAGCTCCTGAA	600
Db	547	GCCTTCAGGAAGAGACAGACAGTTACATGTATAAGAACAGGAATTTGCATCAGCTACTAAA	606
Qy	601	GAATGAGAAAGATGAGGTACAAAAATTTACAAAAATATCATAGCCAGCGCGGCTACTCAGTA	660
Db	607	GAATGAGAAAGATGAGGTGCAAAAAATTTACAAAAATATCATTTGCAAGTCGAGCTACTCAGTA	666
Qy	661	TAATCATGATGCTGAAGAGGAGGAGCGTGAATATAATAGCTTAAAGAGCGCCTGCATCA	720
Db	667	TAATCATGATGAGAGAGAAAGCGGTGAATATAATATAACTTAAGAGAAAGCTCTACATCA	726

1801 GGTTCGAAGAAAGCAGGATCAGAGTGAGCGTGCAGTCGAGCGCCAGCTCGCGGGA 1860  
1804 GGTTCGAGGAGATGTACAAATCAAAATGAGGTGTGCGGTCAAGACTGATCAGCAGGA 1863  
1861 GGGGTGCTACAGCGAGTCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTT 1920  
1864 AGGTGTCTATAGTGGATGCTCTTGGAGCTACACAAATCTCATGTAGAAAAGATGACTT 1923  
1921 ACCTTAATGTGCGGCTGAGTCTGTTCCAGATGTGCGCTAGAGG-----A 1969  
1924 ACCTTAGACATGTGGAGTGGAAATTTTTTCAATTAATGTGTTTCATCAAGTTTCAATCTAA 1983  
1970 GTTCACACAGCGTGTAGCATAAAGTCAGTCGTC-----TAACTTAAGATGCTCAGAGTTG 2024  
1984 GTTGAACACAGGCTGTGTCATTAAGTCAGTTATCTTAATACTTAAGATGCTCAGAGTTG 2043  
2025 TTTGTTTGGACTTCGCTGTCTTCCCCCAAGAGCTGAAATGCT-AAAGCTACTTAAAAAGGA 2083  
2044 TTTGTTTGGACTTCCTGCTCTTCCCCCAAGAGTTGAAATCTTAATCTTATTTAAAAAGGA 2103  
2084 TGCAAAGCTTTGG-----TTGTGTGTAGTAACAGAGCCCTGCTCTGTGACTGCGAGGA 2139  
2104 TATAAAGCTTTGGATATGTATTTTATAGTAACAGAGCATCTGTTTCTGTGAATAAAGGA 2163  
2140 ATGCATGSC-GTTTGGATGGAACAGAGCGCTGGAATGA--TTGCCCTCGCAGGTACCG 2196  
2164 ATGTATAGATGTTTGGATGGAACAGAGCATGAGTGTCTTCTTATAGGTATTA 2223  
2197 AGAAGAGCAGCTTTAGGAGCTGGTTCCTGTAACACATTAATTAATTCGTCCTCAAGTGTGGTT 2256  
2224 AAAATAGCAGCTTTAGGAACTGATTTATGTAATGTTTAAATGTTTGTCTCAATATAGTT 2283  
2257 GGCATTTGGAAGTGTAGCTTTACTTGAATGTATGATGATGATTTTAAACAAGCAGGTTTC 2316  
2284 GGCATTTGGAAGTGTAGCTTTACTTGAATGTATGATGATGATTTTAAACAAGCAGGTTTC 2343  
2317 TATATTTATATGTTTGTAGTGTGATTTTGGGATTAACCTCTTTCATATGTTTT----- 2367  
2344 TATATTTATATGTTTGTAGTGTG-GTTTGAATTAACCTCTTTCATATGTTTAAATAAGT 2402  
2368 -----GTGCTGTACATAAATATACATGACTATGTTTAAAGAGCTTTTAAAGTTT 2415  
2403 GAAATTTATGATGTTTGTACATAGATACATGATTTATGTAAGAGCTTTTAAAGATTT 2462  
2416 AAAAAGCTTACACATGCTTGCATGATGATGATTTTATGCAATTAATGTTTTCAGTGGC 2475  
2463 AAAAGTTTACCA-CAACCATTAAGTATGATTTTATGATGCTTTCATGCTTTCAGTGGT 2520  
2476 ATGTTGTTTACAGAGG--TTAGGACCACTGCGCATGACAGTTAAGACTTTTATTTTAAAG 2533  
2521 ATTCTGTTTACAGATGATTTAGGACCATTTGATGATTTTAAAGAA-ATTCTCTTAAAT 2579  
2534 CCATCTGGGCAATAAAATTCAAAGCCCTTCAT-AAAGTGAATTC---AGATAACTAGA 2589  
2580 ACATCTGGGCAATAAATTTGAAAGGTATTCATGAAAGCTGAGTCTTTAGATTAATCAAC 2639  
2590 ACTACTACGTTACATTTTTCAGATTTTAAAGCAATTTGATTTTATTTATATATGCT-G 2647  
2640 ACTACTAACATTTATATTTTTCAGATTTTATGACATTTAGATTTTATTTTGTATATGTAG 2699  
2648 AATGTTATATTTTCTAAGAGGAATTTGATTTATGAGTAATGAGG 2692  
2700 AATATTAATTTTAAAGGACTATTGATGATAGAGATAGG 2744

RESULT 7  
LOCUS HSM801317 5658 bp mRNA linear HTC 22-SEP-2004  
DEFINITION Homo sapiens mRNA; cDNA DKFZp434C0515 (from clone DKFZp434C0515).  
ACCESSION AL133046  
VERSION AL133046.1 GI:6453458  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 5658)  
AUTHORS Bloecker,H., Boecker,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,  
Osanger,A., Fobo,G., Han,M. and Wiemann,S.  
CONSRM The German cDNA Consortium  
TITLE Direct Submission  
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by GfP (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFZp434C0515) is available at the RZPD Deutsches  
Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp434C0515  
Further information about the clone and the sequencing project is  
available at http://mips.gsf.de/projects/cdna/.  
FEATURES  
Location/Qualifiers  
source  
1..5658  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp434C0515"  
/db\_xref="taxon:9606"  
/clone="DKFZp434C0515"  
/tissue\_type="testis"  
/clone\_lib="434 (synonym: htes3). Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
/note="synovial sarcoma, X breakpoint 2 interacting  
protein"  
gene  
1..5658  
/genes="DKFZp434C0515"  
267..2111  
/genes="DKFZp434C0515"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAB61373.1"  
/db\_xref="GI:6453459"  
/db\_xref="UniProtKB/TrEMBL:Q9UIX0"  
/translation="MGDMVTDPGLSSSEKTSIQVTSKTPSSLYSOVLCSSTIP  
LSKNVSPFSAFTEDNIEQSISYLDQELTTFGPSLYESKGETKREINIVAVLNC  
MNSLLVLQRKNLLAQENVEYONLKLGSMDHLQSCYSKLEQLETSRREMIGLERDR  
KKDKIAMDILNLYVGRADGRKSGWRTKTEARNEDEMYKILLNDYERQOILMENAE  
LKVLQOMKKEMISLLSPQKKPRVDDSTGTVISDVEEDAGELSMWDLSCETV  
REOLTNSIRKOWRILKSHVEKLDONQSVKHLEGNEDVTSRODHEQETKLELEIQ  
CKEMITQOOLLQOOLATVDDDTSLLRDCVYLLEKERLKEWMLPKEOKNPERER  
RSTFEAIRGLERKFAFEERASWLKQOFLNMTTFHONSENVKLFSAFGSSDMDNL  
IVHSRQPKKPSVSNVSGSPVSKLTKSLPASPTSDFCQTRCISSEHSINVLNITA  
EEIKPNQVGGERTNQKWSVASRPSGEGCYSGCSLSYTNSHVEKDDLP"

Query Match 63.2%; Score 1701.4; DB 6; Length 5658;  
Best Local Similarity 80.7%; Pred. No. 0;  
Matches 2214; Conservative 0; Mismatches 471; Indels 60; Gaps 17;  
Qy 1 CGTAGGAGGTGACAGGAGCTGTTGTAAGCGTGCAGCACTGAGCCGCTCTCAGGTAT 60  
Db 188 CATAGAAGAGTGACAGAGCTAGACTAAATGTTTAACTGCTGAAGTATCTCAGGTAT 247  
Qy 61 CTGGCTCTGAACTTCTGATGGAGATTGGAGTCTGTGACAGATCAGTCTCTGTGTAC 120  
Db 248 CTGGCTCTAGAGATTCTGATGGAGATTGGAGTCTGTGACAGTCTGTGCTCTTC 307  
Qy 121 AGAAACAAAAATCTCTCAATATACCTCAGAAACAAAGATGTCTCCGTCAGTTGTA 180  
Db 308 AGAAAGCAAAACTATCTCTCAATATACCTCAGAAACAAAGATGTCTCAATCAAGTTATA 367

QY 181 CTCCAGCAAGTTCTGTGCTCTTCACTACCTTTATCCAAAACGTCGATGGTGTTCGCG 240  
DB 368 CTCACAGCAAGTCTGATGTTCTTCAATACCTTTATCGAAAAATGGCACAGTTTTTTCAG 427  
QY 241 TGTCTTCTGCACAGGAGAAACATTTGAAACAAAGTATTTCTCTATCTTGTATCAGGAGCTGAC 300  
DB 428 TGCCTTCTGCACAGAGATAATATTGAACAGAGTATCTCATATCTTGTATCAGGAATTGAC 487  
QY 301 CACCTTCGGGTTTCTTCTGTTGTATGAAGATTCCAAAGTAAAGAGGCAAGAGAAATT 360  
DB 488 TACTTTTGGTTTCTCTTCAATATATGAAGAAATCCAAAGGTAAGAGACAAAGAGAGAGTT 547  
QY 361 AAATATAGTCGCTGTTCTTGAACCTGTATGAACGAGCTGCTCGTGCTTTCAGCGAAGAACT 420  
DB 548 AAATATAGTAGCTGTAATAATTGATGAGCTGCTTGTGCTTTCAGCGGAAGAACT 607  
QY 421 GCTGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGCGAGTGACATGGACCACT 480  
DB 608 TCTAGCTCAGGAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAAGTATGGACCATCT 667  
QY 481 GCAGAGCTGTACGCCAAACTTTAAGGAGAGTTGGAAACGCTCCAGCGGGAGATGATCGG 540  
DB 668 ACAGAGCTGTACTCAAAACTTTAAGGAAACAACTGGAAAACCTCCAGGAGGGGAAATGATTGG 727  
QY 541 GCTTCAAGAGAGACAGCGAGCTGCAGTGCAGAACAGAGAGTTTGGCATCAGCTCTGAA 600  
DB 728 GCTTCAGGAAAGAGACAGACAGATTACAATGTGAAGAACAGGAATTTGCAATCAGTACTAAA 787  
QY 601 GAAATGAGAAAGTAGAGGTACAAAAATACAAAATATCATAGCCAGCGGGCTACTCAGTA 660  
DB 788 GAAATGAGAAAGTAGAGGTGCAAAAATACAAAATATCATTCGAAGTCGAGCTACTCAGTA 847  
QY 661 TAATCATGATGTGAAGAGAGAGGAGCGTGAATATATAAGCTTAAAGGAGCGCTGCATCA 720  
DB 848 TAATCATGATGTGAAGAGAGAGGAGCGTGAATATATAAGCTTAAAGGAGAACTGATCA 907  
QY 721 GCTCGTTATGAACAAGAGGATAAAAAATAGCCATGATGTTTAAATTTATGCGGTGCG 780  
DB 908 ACTTGTATGAACAAGAAAGATAAGAAATAGCTATGAGACATTTTGAATTTATGCGGGAG 967  
QY 781 AGCTGATGGCAAAACGAGGCTCATGGAGACTGACAAAAACAGAAAGCCAGGAATGAAGATGA 840  
DB 968 AGCTGATGGAAAAAGAGGCTCTCGAGGACTGTTAAACTGAAGCCAGGATGAAGATGA 1027  
QY 841 GATGTACAAAATTTCTGTTGAATGATTATGAGTACCGCCAGAGAGATCTCTGATGGAGAA 900  
DB 1028 AATGTATAAAATTTCTTTGAATGATTATGAATATCGTCAGAAAACAAATCTTAATGGAAA 1087  
QY 901 CGCGAGCTGAAGAGGTCCTCCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCC 960  
DB 1088 TGCAGAACTTGAAGAGGTTCTTCAACAAATGAAAAGGAAATGATTTCTCTCTCTCTCC 1147  
QY 961 TCAGAAGAGAAAGCCAGGAAAGAGCAGAGACGGCACAGGCACTGTTGCTATCTCCGA 1020  
DB 1148 CCAAAAGAGAAACCTAGAGAAAGTAGATGATAGTACAGGAACTGT---TATTTCCGA 1204  
QY 1021 TATAGAAGATGATCTTGGGGAACCTAGCAGAGACAGCGGTGTGGGCGCTTTCTGTGACAC 1080  
DB 1205 TGTGAAGAGATGCGGGGGAACCTAAGCAGAGAGATGATGTGGGACCTTTCTCTGTAAC 1264  
QY 1081 TGTGAGAGAGAGCTGACAAACAGCATCCAGGAACAGTGGAGAAATTTTGAAGAGTCATGT 1140  
DB 1265 TGTGAGAGAGAGCTTACAAACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCATGT 1324  
QY 1141 AGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGGCCCTTAATGAGGAGGAGCT 1200  
DB 1325 AGAAAAGCTTGTAAACCAAGTTTCAAGGTTACACTGGAAGGTTTAAATGATGAAGATGT 1384  
QY 1201 CATCTCAGCAAGACCATAGCAAGAGAGATGAGAAACTGAGAGCTGGAGATTGAGCGGTG 1260  
DB 1385 AATCTCAGCAAGACCATGAACAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTG 1444  
QY 1261 TANAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCAGCTGTGA 1320

DB 1445 TAAAGAAATGATTAATAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTCATATGA 1504  
QY 1321 TGATGACACCACTCCTCAGTGTTCGAGAGACTGTTTACTTGTGGAAGAAAAGGAACGCTTAA 1380  
DB 1505 TGATGATACCACTTCACTATTACGAGACTGTTTATTTGTTGGNAGAAAAGGAACGCTCAA 1564  
QY 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAGAGAGAAAAGCGAAGCTT 1440  
DB 1565 AGAAGATGGTCCCTTTTAAAGAGCAGAAAAGAAATTTTGAGAGGGAGAGACGAAGCTT 1624  
QY 1441 TACAGAACTGCCATTCGATTGGGTTTGGAGAGAAAGGGTTTGAAGAGACGCGCCAG 1500  
DB 1625 TACAGAGCCGCTATTTCGCCTGGGATTGGAGAGAAAAGGCAATTTGAAGAGAAAGAGCCAG 1684  
QY 1501 CTGGTAAAGCAGCAGATTTTTAAACATGACGAACCTTTGACCAACCAAGAACTCAGAAAAATGT 1560  
DB 1685 TTGGTTAAAGCAGCAGATTTCTAATATGACTTACCTTTGACCAACCAAGAACTCAGAAAAATGT 1744  
QY 1561 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTCACTCGAG 1620  
DB 1745 GAAACTTTTTCAGTGCCTTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTCACTCGAG 1804  
QY 1621 GCCAGCGCAAAAGAGCTACACAGTGTGGCTAATGGGGTGCAGCTTGCACATCAAAACT 1680  
DB 1805 GCAGCCGCAAAAGAGCGCTCAAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTAAACT 1864  
QY 1681 GACTAAATCTCTTCTGCTCCACCTTCTACTTTCAGACTTTCCGCCAGACACATTTCAATGTGT 1740  
DB 1865 TACTAAATCTCTTCTGCTTCCACCTTCCACTTTCAGACTTTTCCGCCAGACACGTTTCTCGAT 1924  
QY 1741 GTCTGAACACAGTTCCATCAGTGTGCTGAATATATACTCTGAAAGAAAGTAAACCAAGTGA 1800  
DB 1925 ATCTGAACATAGTTCAATCAATGACTGAAATATACTGCTGAAGAAATTAACCAAAATCA 1984  
QY 1801 GGTTCGAAGAGAAAGCAGGATCAGAACTGAGAGCTGAGCTGAGGCGCCAGCTCCGCGGA 1860  
DB 1985 GGTTCGAGGAGAAAGCTACAAATCAAAATGAGGTGTGGCGTCAAGACCTGGATCAAGGA 2044  
QY 1861 GGGGTGCTTACAGCGGATGCTCCTCGGCTTTCAGAGAGCGCTCACGGGGACCCAGAGTACTT 1920  
DB 2045 AGGTTGCTATAGTGAATGCTCTTGAAGTACACAAATCTCATGTAGAAAAGAGTACTT 2104  
QY 1921 ACCTTAAATGTGCGGCTGAGTGTGCTGCCAGATGTGCGCTAGAGG-----A 1969  
DB 2105 ACCTTAGACATGTGGAAGTGGAAATTTTTCATTAATGTGTTTCATCAAGTTTTCACATCTAA 2164  
QY 1970 GTTCACACAGGCTGATGATTAAGTCACTGCTC-----TAACTTAAAGTGTCTCAGAGTTG 2024  
DB 2165 GTTGAACAGGGTGTGTCATAAAGTCACTTCTTAATAAATTAAGATGGTCTGAGTTG 2224  
QY 2025 TTTGTTTGGACTTCGCTGCTTCCGCCAAAGAGCTGAAATGCT--AAGCTACTTAAAGGA 2083  
DB 2225 TTTGTTTGGACTTCGCTGCTTCCGCCAAAGAGTTGAATCTTAAATCTATTATAAGGA 2284  
QY 2084 TGCAAAAGCTTTGG-----TTGTGTGTAGTAACAGAAGCCCTGCTCTGTGACTGCAAGGA 2139  
DB 2285 TATAAAGCTTTGGATATGATTTTATAGTAACAGAAGCATCTGGTTCTGTGAATAAGGA 2344  
QY 2140 ATGCATGCG--GTTTGGATGNAACAGAGCGCTGGAATGA--TTGCGCTCGCCAGGTACCG 2196  
DB 2345 ATGTATAGATGTTTGGATGNAACAAAGCAGTACAGTGAAGTTTCTCTTATATAGGTATTA 2404  
QY 2197 AGAAGAGCACTTTTAGGGAGCTGGTTCTGTAAACATTTAAATATTTCTGCTCCCAAGTGTGTT 2256  
DB 2405 AATAAGCACTTTTAGGAAACTGATTAATGTTAAATGTTTGTCTCAATATAGTT 2464  
QY 2257 GGCAATTGGAAGTGTAGCCTTTACTTGAATGTATCTAGTATGATTTTAAACAAAGCAGGTTT 2316  
DB 2465 GGCATTGGAAGTTTAGCCTTTACTTGAATGTATCTAGTATGATTTTAAACAAAGCAGGTTT 2524  
QY 2317 TATATTTTATGTTTGTAGTGTATTTGGGATTAACCTTCTTTCATATGTTTT----- 2367

Db	2525	TATATTTATGTTTATGTTG - GTTTGAAATACCTCTTTTCATATGCTTTTAAATAAAGT	2583
Qy	2368	-----GTGCTGTACATAAATATACATGACTATGTTTAAAGAGGCTTTTAAAGCTTT	2415
Db	2584	GAATTTATGATGTTTGTACATAGATACACATGATTATGTTTAAAGAGGCTTTAAGATTT	2643
Qy	2416	AAAAAATTCACACCATGCTTTGAGTATAGCATTTTCATGCCAATTAATAATGTTTTCAGTGGC	2475
Db	2644	AAAAAGTTTCACA - CAACCATAGATATAGTATTTTCATGCC - AGTAAAAATTTTTTAGTGGT	2701
Qy	2476	ATGCTGTTTACAGAGG--TTAGGACCACTGCCACATGACATGACAGTTTAAAGACTTTTATTTAAAG	2533
Db	2702	ATTCGTTTACAGATGATTAGGACCATGATGCAATTACATTTAAGA - ATTCTCTTTTAAT	2760
Qy	2534	CCATCTGGGCAATAAAAAATTCAAAGCCCCCTTCAT - AAGCTGAGTTC--AGATAACTAGA	2589
Db	2761	ACATCTGGGCAATAAATATTGAAAGGTATTCCATGAAGCTGAGTCTTTTATAGATAATCAAC	2820
Qy	2590	ACTACTAAGCTTACATTTTGGAGATTTTAAAGCAAT - GTATTTTATTTTATATATAGT - G	2647
Db	2821	ACTACTAATCATTTTGGAGATTTTATGACATTTAGATTTTATTTTGTATATAGTAG	2880
Qy	2648	AATGTTATAATTTCTAAGAGGAATATTGATTTATGGAGTAATGGG	2692
Db	2881	AATATTATAATTTTAAAGGACTATTGATGATAGAAGAAATAGG	2925
RESULT 8			
CR613292			
LOCUS			
DEFINITION	CR613292 2111 bp mRNA linear HTC 21-JUL-2004 full-length cDNA clone CS0DL007Y108 of B cells (Ramos cell line)		
ACCESSION	CR613292		
VERSION	CR613292.1 GI:50494099		
KEYWORDS	HTC; CNSLT cDNA.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2111)		
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.		
JOURNAL	Full-length cDNA libraries and normalization		
REMARK	Unpublished		
REFERENCE	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue		
AUTHORS	2 (bases 1 to 2111)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
COMMENT	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
FEATURES	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR v sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.		
source	Location/Qualifiers		
	1. 2111		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CS0DL007Y108"		
	/tissue_type="B cells (Ramos cell line) Cot 25-normalized"		
	/plasmid="pCMVSPORT_6"		
ORIGIN			
Query Match	52.6%; Score 1416.6; DB 6; Length 2111;		
Best Local Similarity	83.1%; Pred. No. 0;		
Matches 1627; Conservative	0; Mismatches 329; Indels 3; Gaps 1;		
Qy	1	CGTAGGAGAGTGACAGGAGCTGTTGTAAGCGTCGACGACTGAGCCGCTCCTCAGGTAT	60

Db	127	CATAGAAGAGTGACACAGCTAGACTAAATGTTTAACTGCTGAACCTAGTTTCCCTCAGGTAT	186
Qy	61	CTGGCTCTGGAACTTGTCTATGGGAGATTGGATGACGTGACAGATCCAGTTCTGTGTAC	120
Db	187	CTGGCTCTAGAGATTGCTATGGGAGATTGGATGACTGTTTACAGATCCAGGTCTGTCTTC	246
Qy	121	AGAAACAAAAATCTCTCAATATACCTCAGAAAAAAGATGTCCTCGTCCAGTTTGTGA	180
Db	247	AGAAAGCAAAACTATCTCTCAATATACCTCAGAAAAAAGATGTCCTCCATCAAGTTTATA	306
Qy	181	CTCCAGCAAGTTCTGTGCTCTTTCAGTACCTTTTATCCAAAAAGCTGTCATGGTTCGG	240
Db	307	CTCAGAGCAAGTCTATGTTCTTCAATACCTTTATCGAAAAATGTCACAGTTTTCAG	366
Qy	241	TGCTCTTCGACAGGAGAGAACATTGAAACAAAGTATTTCTCTATCTTGATCAGAGCTGAC	300
Db	367	TGCTTCTGACAGAAAGATAATATTGAAACAGAGTATCTCATATCTTGTGATCAGGAATTC	426
Qy	301	CACCTTCGGGTTTCCCTTCTGATGAAGAATCCAAAAAGTAAAGAGGCAAGAGAGAAAT	360
Db	427	TACTTTTGGTTTCTCTTCAATATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGATT	486
Qy	361	AAATATATGCTGTTCTGAACTGTATGAACGAGCTGCTCGTCTCTCAGCGGAAGAACCT	420
Db	487	AAATATAGTAGCTGTACTAAATTTGTTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCT	546
Qy	421	GCTGGCCAGGAGAGCGTGGAGACACAGAACTTTGAAGCTGGGAGTGACATGACCACCT	480
Db	547	TCTAGCTCAGGAAAAATTTGTGGAGACACAGAAATTTGAAGCTGGGAAGTATATGACCCTCT	606
Qy	481	GCAGAGCTGCTACGCCAAACTTAAAGAGCAGTTTGGAAACGTCAGCGCGGAGATGATCGG	540
Db	607	ACAGAGCTGCTCTCAAACTTTAAGGAACAACTGGAACAACTCCAGGAGGGAATGATTGG	666
Qy	541	GCTTCAAGAGAGACAGCAGCGCTGCAGTGCAGAAACAGAGAGTTTGCATCAGCTCTGAA	600
Db	667	GCITTCAGGAAAGAGACAGACAGTTCACATGTAAGAACAGGAAATTTGCATCAGCTACTAAA	726
Qy	601	GAATGAGAAAGATGAGTACAAAAATTTACAAAAATATCATAGCCAGCCGGCTACTCAGTA	660
Db	727	GAATGAGAAAGATGAGTGCAAAAATTTACAAAAATATCATTTGCAAGTCGAGCTACTCAGTA	786
Qy	661	TAATCATGATGTCAGAGGAGGAGCGTGAATATAAAGCTAAAGAGGCGCTCTGCAATCA	720
Db	787	TAATCATGATGTCAGAGGAAAGAGCGTGAATATAAAGCTGAAGGAACTCTACATCA	846
Qy	721	GCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGTTTTAAATTTATGTGGTCTG	780
Db	847	ACTTGTATTGAACAAGAGATTAAGAAAAATAGCTATGGACATTTTGAATTTATGTGCGGAG	906
Qy	781	AGCTGATGCCAAACGAGGCTCATGGAGGACTGCACAAACAGAGCCAGGAATGAAGATGA	840
Db	907	AGCTGATGGAAAAAGAGGCTCTCGGAGGACTGTTAAACCTGGAAGCCAGGAATGAAGATGA	966
Qy	841	GATGTACAAAAATTCGTTGTAATGATTATAGTAGTACCGCCAGAACAGACATCTGTATGGAGAA	900
Db	967	AATGTATAAAATTCCTTGTGATGATTATGATATCGTCAGAACAAATCTCTATGGAAAA	1026
Qy	901	CGCGGAGCTGAAGAAGGTCTCTCAGCAGATGAAGAAGAGATGATCTCTCTCTCTCTCC	960
Db	1027	TGCAGAACTTAAAGAGGTTCTTCAAACAAATGAAAAAGGAAATGATTTCTCTCTCTCTCC	1086
Qy	961	TCAGAGAGAGAACCCAGGAGAGAGAGAGGAGCGGCACAGGCACTGTTGCTATCTCCGA	1020
Db	1087	CCAAAAAGAGAAACCTAGAGAAAAAGAGTAGATAGTATCAGGAACTGT---TATTTCCGA	1143
Qy	1021	TATAGAAGATGACTCTGGGAACTGTAGCAGAGACAGCGTGTGGGCGCTTTCTGTGTACAC	1080
Db	1144	TGTTGAAGAGATGCCGGGAACTAAGCAGAGAGATGATGTGGACCTTTCTCTGTGAAAC	1203
Qy	1081	TGTGAGAGACAGCTGACAAACAGCATCAGGAAACAGTGAGAAATTTTGAAGAGTCATGT	1140

Db	1204	TGTGAGAGCAGCTTACAAA	CAGCATCAGAAAA	CAGTGGAGAAATTTT	GGAAAGTCATGT	1263
Qy	1141	AGAAAACTCGATACCAAGCTTCGAAGGTACACT	TCAGAGGGCCCTTAATGAGGAGGACGT			1200
Db	1264	AGAAAAGCTTGATAACCAAGTTTCAAAGGTACA	CCTGGAAGTTTAAATGATGAAGATGT			1323
Qy	1201	CATCTCAGCAGAACCCATGACGAAGAGACTGAG	AAACTGGAAGCTGGAGATTGAGCGGTG			1260
Db	1324	AATCTCAGCAGAACCAACCATGAACGAACACT	GAAAACTCGAGTTAGAAATTCAGCAGTG			1383
Qy	1261	TAAAGAGATGATCAAGGCTCAGCAGCAGCTCTT	TACAGCAGCAGCTGCCACCACGTGTGA			1320
Db	1384	TAAAGAAATGATTAAAACTCAGCAACAGCTTT	TACAGCAGCAGCTCGTACTGCTATATGA			1443
Qy	1321	TGATGACACCACTCAGCTTGTTCGAGAGCTGTT	ACTTGTCTGGAAGAAAAGGAAACCCCTTAA			1380
Db	1444	TGATGATACCACTTCACTATTACGAGACTGTT	ATTGTTGGAAGAAAAGGAAACGTCTCAA			1503
Qy	1381	AGAAGAGTGGACCTTTTAAAGAGCAAAAAAG	AAATTTTGAGAGAGAAAGCGGAGCGT			1440
Db	1504	AGAGAAATGGTCCCTTTTAAAGAGCAGAAAA	AGAAATTTTGAGAGGGGAGAGACGAAGCTT			1563
Qy	1441	TACAGAAGCTGCCAATTCGATTTGGGTTGGAG	AGAAAGCGTTTGAAGAGAGCGAGCCAG			1500
Db	1564	TACAGAAGCGCTATTTCGCCCTGGGATTTGG	AGAGAAAGGCCATTTTGAAGAAAGAAAGAG			1623
Qy	1501	CTGGTAAAGCAGCAGTGTTTTAAACATGAC	GAACTTTTGACCACACAGAACTCAGAAAA			1560
Db	1624	TTGGTTAAAGCAGCAGTTTCTAATATGACTA	CTCTTTTGACCAACAGAACTCAGAAAA			1683
Qy	1561	GAAACTTTTCAGTGCTCTCAGGAAGTTCTG	ATTGGGACAACTTTATAGTCCACTCAG			1620
Db	1684	GAAACTTTTCAGTGCTCTCAGGAAGTTCTG	ATTGGGACAACTTTATAGTGCACCTCG			1743
Qy	1621	GCCAGGCAAAAGAGCTACACAGTGTGGCT	TAATGGGGTGCAGCTTGACATCAAA			1680
Db	1744	GCAGCGCAAAAGAAAGCCCTCACAGTGTGT	CTAATGGGTCTCCAAGTTTGCATGTCT			1803
Qy	1681	GACTAAATCTCTTCTCGCTCACCCTTCTAC	TTCCAGACTTTTGCAGACACATTCATGT			1740
Db	1804	TACTAAATCTCTTCTCGCTTCACTTCCACT	TTTGCAGACACAGTTCTCTGCAAT			1863
Qy	1741	GTCGAAACACAGTTCATCAGTGTGCTGAA	TATAACTCTCTGAAAGAAAGTAAACCA			1800
Db	1864	ATCTGAACATAGTTCATCAATCAATGACT	GATAATAACTGCTGAGAAATTAACCA			1923
Qy	1801	GGTTGCAAGAGAAAGCAGGATCAGAAGT	GCAGGTCAGTGCAGGCCAGCTCGCGG			1860
Db	1924	GGTTGGAGAGAAATGTACAAATCAAA	ATGAGTGTGGCATCAAGACTCGATCAG			1983
Qy	1861	GGGTGCTCAGCGGATGCTCTCGGCTTT	CAGAGCGCTCACGGGACCGAGATGACT			1920
Db	1984	AGGTTGCTATGTGGATGCTCTCTTGAGCT	ACACAAATTCATGTAGAAAAATGACT			2043
Qy	1921	ACCTTAAATGTCGGGCTGCAGTGTCTG	TCTCCAGATGTG			1959
Db	2044	ACCTTAGACATGTGGACTGGAAATTTT	TTTTCATTAATGTG			2082
RESULT 9						
AY412491						
LOCUS	1845 bp DNA linear GSS 12-DEC-2003					
DEFINITION	Homo sapiens HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.					
ACCESSION	AY412491					
VERSION	AY412491.1					
KEYWORDS	GI:39768456					
SOURCE	GSS.					
ORGANISM	Homo sapiens (human)					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 1845)					

AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.					
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios					
JOURNAL	Science	302	(5652),	1960-1963	(2003)	
PUBMED	14671302					
REFERENCE	2	(bases 1 to 1845)				
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.					
TITLE	Direct Submission					
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA					
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.					
FEATURES	Location/Qualifiers					
source	1..1845					
	/organism="Homo sapiens"					
	/mol_type="genomic DNA"					
gene	/db_xref="taxon:9606"					
	<1..>1845					
	/locus_tag="HCM4575"					
ORIGIN						
Query Match	50.7%; Score 1363.8; DB 14; Length 1845;					
Best Local Similarity	84.0%; Pred. No. 0;					
Matches 1552; Conservative	0; Mismatches 292; Indels 3; Gaps 1;					
Qy	80	ATGGGAGATTGGATGACTGTGACAGATCCAGTCTCGTGTACAGAAAAACA	AAATCTCTCT	139		
Db	1	ATGGGAGATTGGATGACTGTGTACAGATCCAGTCTGTCTTCAGAAAGCA	AAATCTCTCT	60		
Qy	140	CAATATACCTCAGAAAAACAAGATGTCTCCGTCCAGTTTGTACTCCAGCA	AGTTCTGTGC	199		
Db	61	CAATATACCTCAGAAAAACAAGATGTCTCCATCAAGTTTATACTCACACA	AGTTGTATGT	120		
Qy	200	TCCTTCAGTACCTTTATCCAAAAAGTGCATCGTGGTGTCTTCCGGTGTCT	CTGACAGAGAG	259		
Db	121	TCCTTCATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGGCTT	CTGACAGAGAT	180		
Qy	260	AACATTGAAACAAAGTATTTCTCTATCTTGATCAGAGCTGACCATCTCG	GGGTTTCTCTCC	319		
Db	181	AATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTACTTTTG	GGTTTTCTCTTCA	240		
Qy	320	TTGTATGAGAAATCCAAAAGTAAAGAGGCAAGAGAGAAATTAATATAGT	CGCTGTCTGTG	379		
Db	241	TTATATGAAGAATCCAAAGGTAAAGACAAAGAGAGAGTTAAATATAGT	ACTGTACTACTA	300		
Qy	380	AACCTGTATGAACGAGCTGCTCGTCTTCAGCGGAAGAACCTCTGGCC	CCAGGAGAGCGTG	439		
Db	301	AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCT	CAGGAAATGTG	360		
Qy	440	GAGACACAGAACTTGAAGCTGGGAGTGCATGACATGACCACTTCGAGAG	CTGCTACGCCAAA	499		
Db	361	GAGACACAGAAATTTGAAGCTGGGAAGTGAATGGAACCATCTACAGAG	CTGCTACTCAAAA	420		
Qy	500	CTTAAGGAGCAGTTGGAAAAGTCCAGCGGGAGATGATCGGGCTTCAAG	AGAGAGACAGG	559		
Db	421	CTTAAGGAAACAACCTGGGAAACCTCCAGGAGGGAATGATTGGGCTTC	AGGAAAGACAGA	480		
Qy	560	CAGCTGCAGTGCAGAAACAGAGTTTGCATCAGCTCTCTGAAAGAT	TGAGAAAGATGAGGTA	619		
Db	481	CAGTTACAAATGTGAAGAACAGGAATTTGCATCAGCTACTAAAGAAT	TGAGAAAGATGAGGTG	540		
Qy	620	CAAAAATTACAAAATATCATAGCCAGCGGGCTACTCAGTATATCATG	TGATGTGAAGAGG	679		
Db	541	CAAAAATTACAAAATATCATTTGAAGTCGAGCTACTCAGTATATCAT	GATGATGAAGAGA	600		
Qy	680	AAGGAGCGTGAATATATAAGCTTAAAGGAGCGCTGCAATCAGCTCGT	TATGAACAAGAG	739		

Db 601 AAGAGCGTGAATATAATAAAGCAAGCAAGTCTCATCAACTGTTGTTATGAACAAGAA 660  
Qy 740 GATAAAAACATAGCCATGGATGTTTTTAAATATATGTGGTTCAGCTGATGGCAACAGAGGC 799  
Db 661 GATAGAAAAATAGCTATGGACATTTTGAATATATGTCTGGGAGAGCTGATGGAAAAAGAGGC 720  
Qy 800 TCATGGAGGACTGACAAAAACAGAACAGGCAAGTCAAGATGAGATGTACAAATTTCTGTG 859  
Db 721 TCCTGGAGGACTGGTAAACTGAAGCAGGAAATGAAGATGAATGTATAAATTTCTCTTG 780  
Qy 860 AATGATTATAGTACCGCAGAGCAGATCCTGATGGAGAACGCGGAGCTGAAGAAGTC 919  
Db 781 AATGATTATGAATATCGTCAGAAAAAATCTAATGGAATAATGCAGAACTTAAGAAGTT 840  
Qy 920 CTCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 979  
Db 841 CTTCAACAAATGAANAAGGAATGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Qy 980 GAAAGAGCAGAGGAGCGGCACAGGCACTGTTCTCTATCTCCGATATAGAAAGATGACTCTGGG 1039  
Db 901 GAAAGAGTATGATGATGACAGGAAGTGT---TATTTCCGATGTTGAAGAAGATGCCGGG 957  
Qy 1040 GAACGTGAGCAGAGCAGCGTGTGGGGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1099  
Db 958 GAACTAAGCAGAGAGAGTATGTGGGACCTTTCTCTGTGAACCTGTGAGAGAGCAGCTTACA 1017  
Qy 1100 AACAGCATCAGGAACAGTGGAGAAATTTTGAAGTCAATGTAGAAAAAAGTCTGATAACCAA 1159  
Db 1018 AACAGCATCAGAAACAGTGGAGAAATTTTGAAGTCAATGTAGAAAAAGTCTGTAAACCAA 1077  
Qy 1160 GCTTCGAAGGTACACTCAGAGGGCTTAATCAGAGGAGCGTCATCTCAGCAACAGACCAT 1219  
Db 1078 GTTTCAAGGTACACTCGGAAGTTTAAATGATGAAGATGTAACTCAGCAACAGACAT 1137  
Qy 1220 GAGCAAGAGACTGAGAACTGAGACTGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1279  
Db 1138 GAACAAGAACTGAAAAAAGTCTGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAACT 1197  
Qy 1280 CAGCAGAGCTTTACAGCAGCAGCTGGCCACACAGCTGTGATGATGATCACACACCTCACTG 1339  
Db 1198 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATACACCTTCACTA 1257  
Qy 1340 TTGGCAGACTGTTACTTCTGGAAGAAAGAAAGCGCTTAAAGAGAGTGGACCCCTTTT 1399  
Db 1258 TTACGAGACTGTTATTTGTCGAGAAAGAAAGCGTCTCAAGAAAGATGTCCTCTTTT 1317  
Qy 1400 AAAGAGCAAAAAAGAAATTTTGAGAGAAAGGCGAAGCTTTTACAGAGCTGCCATTCGA 1459  
Db 1318 AAAGAGCAAAAAAGAAATTTTGAGAGGAGAGAGCAAGCTTTTACAGAGCGCGCTATTTCG 1377  
Qy 1460 TTGGGGTTGGAGAGAAAGGCGTTTGAAGAGAGCGGCGCAGCTGGGTAAGCAGCAGTTT 1519  
Db 1378 CTGGGATTTGGAGAGAAAGGCAATTTGAAGAGAAAGAGCGCAGTTGGTTAAAGCAGCAGTTT 1437  
Qy 1520 TTAACATGACGCAACTTTGACCAACAGAACTCAGAAAAATGTGAAACTTTTTCAGTGCCTTC 1579  
Db 1438 CTAATATGACTACTTTTGACCAACAGAACTCAGAAAAATGTGAAACTTTTTCAGTGCCTTC 1497  
Qy 1580 TCAGAAAGTTCTGATCCAGAACTTTATAGTCACTCA CGGCCAGCGCAAAAAAGAGCTA 1639  
Db 1498 TCAGAAAGTTCTGATTTGGGACAACTTTATAGTGCACTCGAGGCGAGCGCGCAAAAAAGAGCCT 1557  
Qy 1640 CACAGTGTGGTAAATGGGTGCGAGCTTGACATCNAACCTGACTTAATCTCTTCCTGCC 1699  
Db 1558 CACAGTGTGTCTAATGGGTCTCCAGTTTGGATGCTTAAACTTACTTAAATCTCTTCCTGCT 1617  
Qy 1700 TCACCTTCTACTTTCAGACTTTTCGCGACACAACTCATGTGTGTCTGAAACAGTTTCCATC 1759  
Db 1618 TCACCTTCCACTTCAGACTTTTTCGCGACACAGCTTCTCTGCACTATCTGAACATAGTTCAATC 1677  
Qy 1760 AGTGTGCTGAATAAATCTCTGAGAAAGTAAACCAAGTGAAGTTGCAAGAGAAAGCAG 1819  
Db 1678 AATGTACTGAATATAAATCTGCTGAAGAAATTTAAACCAATTCAGTTTGGAGGAGATGTACA 1737

Qy 1820 GATCAGAGTGGAGCGTGCAGTCCAGGCCAGCTCGCGGAGGGGTCTACAGCGGATGC 1879  
Db 1738 AATCAAAAATGGAGTGTGGCATCAAGACCTGGATCAGAGAAAGTTCTATAGTGGATGC 1797  
Qy 1880 TCCTCGGCCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCTTA 1926  
Db 1798 TCCTTGAGCTACAAATTTCTCATGTAGAAAAAGATGACTTACCTTA 1844  
RESULT 10  
AY412492 1845 bp DNA linear GSS 12-DEC-2003  
LOCUS Pan troglodytes HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,  
DEFINITION genomic survey sequence.  
ACCESSION AY412492  
VERSION AY412492.1 GI:39768457  
KEYWORDS GSS.  
SOURCE Pan troglodytes (chimpanzee)  
ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.  
REFERENCE 1 (bases 1 to 1845)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 1845)  
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
Location/Qualifiers  
source 1..1845  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
gene <1..>1845  
/locus\_tag="HCM4575"  
ORIGIN  
Query Match 50.4%; Score 1357.4; DB 14; Length 1845;  
Best Local Similarity 83.5%; Pred. No. 0;  
Matches 1542; Conservative 0; Mismatches 302; Indels 3; Gaps 1;  
Qy 80 ATGGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTACAGAAAAAATCTCTCT 139  
Db 1 ATGGGAGATTGGATGACTGTGTTACAGATCCAGGCTGTCTTCAGAAAGCAAACTATCTCT 60  
Qy 140 CAATATACCTCAGAAAAAAGATGTCTCCGTCCAGTTTGTACTCCAGCAAGTTCTGTGC 199  
Db 61 CAATATACCTCAGAAAAAAGATGTCTCCATCAAGTTNNNACTCACAGCAAGTCTATGT 120  
Qy 200 TCTTCAGTACTTTATCCAAAAAGCTGCATGGTGTCTTCGGTGTCTCTGTCACAGGAGAG 259  
Db 121 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCAAGAAGAT 180  
Qy 260 AACATTGAACAAAGTATTTCTCTATCTTTGATCAGGAGCTGACCACTTCGGGTTTCTCTCC 319  
Db 181 AATATTGAACAGAGTATCTCATATCTTTGATCAGGAATTGACTACTTTTGGTTTTCTTCA 240  
Qy 320 TTGTATGAAGAAATCCAAAAAGTAAAGAGGCAAGAGAAATTAATATAGTCGCTGTCTG 379

[illegible]



/note=:Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I; Site\_2: Not I; The library was constructed according to Bonaïdo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is NATTAATTAGC. This library was created for the University of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH)."

INSTITUTE OF MEDICAL RESEARCH (NRMH)									
ORIGIN	Query Match Best Local Similarity Matches	32.6%	Score 876.4;	DB 9;	Length 878;				
		99.9%;	Pred. No. 7.9e-215;	Mismatches 0;	Indels 0;	Gaps 0;			
		877;	Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;		
Qy	388	GAACGAGCTCTCGTGTCTTCAGCGGAAGAACCTGCTGCCCCAGAGAGAGCGTGGAGACACA	447						
Db	1	GAACGAGCTGCTCGTGTCTTCAGCGGAAGAACCTGCTGCCCCAGAGAGAGCGTGGAGACACA	60						
Qy	448	GAACCTTGAAGCTGGCGCAGTGACATGSGACCACCTCTCAGAGCTGCTACGCCAAACTTTAAGGA	507						
Db	61	GAACCTTGAAGCTGGCGCAGTGACATGSGACCACCTCTCAGAGCTGCTACGCCAAACTTTAAGGA	120						
Qy	508	GCAGTTGGAAACGTCCAGGCGGGAGATGATCGGCCTTCAGAGAGAGACAGGCGAGCTGCACA	567						
Db	121	GCAGTTGGAAACGTCCAGGCGGGAGATGATCGGCCTTCAGAGAGAGACAGGCGAGCTGCACA	180						
Qy	568	GTGCAAGAACAGGAGTTTGGCATCAGCTCCTCGAAGAAATGAGAAACATCAGAGTACAAAAATT	627						
Db	181	GTGCAAGAACAGGAGTTTGGCATCAGCTCCTCGAAGAAATGAGAAAGATGAGGTACAAAAATT	240						
Qy	628	ACAAAATATCATAGCCAGCCGGGCTACTCAGTATATAATCATGATGTGAAGAGGAAGAGCG	687						
Db	241	ACAAAATATCATAGCCAGCCGGGCTACTCAGTATATAATCATGATGTGAAGAGGAAGAGCG	300						
Qy	688	TGATATAATAAGCTTAAGGAGCCCTGCATCAGCTCGTTATGAACAAGAGAGGATAAAAA	747						
Db	301	TGAATATAATAAGCTTAAGGAGCCCTGCATCAGCTCGTTATGAACAAGAGAGGATAAAAA	360						
Qy	748	CATAGCCATGGATGTTTTAAATTAATGTGGGTCGAGCTGATGGCAACAGAGGCTCATGGAG	807						
Db	361	CATAGCCATGGATGTTTTAAATTAATGTGGGTCGAGCTGATGGCAACAGAGGCTCATGGAG	420						
Qy	808	GACTGACAAACAGAACGCCAGGAATGAAGATGATGATACAAATTTCTGTTGAATGATTA	867						
Db	421	GACTGACAAACAGAACGCCAGGAATGAAGATGATGATACAAATTTCTGTTGAATGATTA	480						
Qy	868	TGAGTACCGCCAGAAAGCAGATCCTGATGGAGAACCGGAGCTGAAGAAGGTCTCCACGACA	927						
Db	481	TGAGTACCGCCAGAAAGCAGATCCTGATGGAGAACCGGAGCTGAAGAAGGTCTCCACGACA	540						
Qy	928	GATGAAGAAGAGATGATCTCTCTCTCTCTCAGAGAAGAACGCCAGGGAAAGAGC	987						
Db	541	GATGAAGAAGAGATGATCTCTCTCTCTCTCAGAGAAGAACGCCAGGGAAAGAGC	600						
Qy	988	AGAGGACGGCACAGGCACTGTTGCTATCTCCGATATAGAGATCACTCTCGGGAACCTGAG	1047						
Db	601	AGAGGACGGCACAGGCACTGTTGCTATCTCCGATATAGAGATCACTCTCGGGAACCTGAG	660						
Qy	1048	CAGAGACAGCGTGTGGGCGCTTTCTCTGTGACACTGTGTGAGAGCAGCTGACAAACAGCAT	1107						
Db	661	CAGAGACAGCGTGTGGGCGCTTTCTCTGTGACACTGTGTGAGAGCAGCTGACAAACAGCAT	720						
Qy	1108	CAGGAACAGTGGAGAAATTTGAAAAGTCATGTAGAAAACCTCAATACCAAGCTTCGAA	1167						
Db	721	CAGGAACAGTGGAGAAATTTGAAAAGTCATGTAGAAAACCTCAATACCAAGCTTCGAA	780						
Qy	1168	GGTACACTCAGAGGGCCCTTAATGAGGAGGAGCGTCACTCTACGACAAGACCATATGACGA	1227						

Db	781	GGTACACTCAGAGGGCGCTTAATGAGGAGGACGTCATCTCAGCAAGACCATGACGA	840
Qy	1228	GACTGAAAACCTGGAGCTGGAGATTGAGCGGTGTAAAG	1265
Db	841	GACTGAAAACCTGGAGCTGGAGATTGAGCGGTGTAAAG	878
RESULT 12	CF949831	UT1-M-H30-cmt-d-22-0-UI.r1 NIH BMAP_H30 Mus musculus	linear EST 20-NOV-2003
LOCUS	CF949831	IMAGE:30632061 5', mRNA sequence..	
DEFINITION	CF949831	CF949831.1 GI:38465700	
ACCESSION	CF949831	CF949831.1	
VERSION	CF949831	CF949831.1	
KEYWORDS	CF949831	CF949831.1	
SOURCE	CF949831	CF949831.1	
ORGANISM	CF949831	CF949831.1	
REFERENCE	CF949831	CF949831.1	
AUTHORS	CF949831	CF949831.1	
TITLE	CF949831	CF949831.1	
JOURNAL	CF949831	CF949831.1	
COMMENT	CF949831	CF949831.1	

ORIGIN	Query Match	30.2%	Score 812;	DB 5;	Length 840;
	Best Local Similarity 99.1%;	Pred. No. 3.5e-198;			
	Matches 836;	Conservative 0;	Mismatches 4;	Indels 4;	Gaps 2;
Qy	1127	TTGAAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAAGGTACACTCAGAGGGCGTT			1186
Db	1	TTGAAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAAGGTACACTCAGAGGGCGTT			60
Qy	1187	AATGAGGAGACGTCATCTCCAGCAAGACCCATGACGACAGAGACTGAGAAACTGGAGCTG			1246



```
Db      61  AATGAGGAGGAGCTCATCTTCAGACACAGACCATGACCAAGAGACTGAGAAACTGGAGCTG 120
Qy      1247  GAGATTGACGGGTGTAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGAGCTG 1306
Db      121  GAGATTGACGGGTGTAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCTG 177
Qy      1307  GCACACACCTGTGATGATGACACACCTCCTCACTGTTGCGAGACTGTTACTTGTGGAAGA 1366
Db      178  GCACACACCTGTGATGATGACACACCTCCTCACTGTTGCGAGACTGTTACTTGTGGAAGA 237
Qy      1367  AAGGAACGCGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAGAAATTTGAGAGA 1426
Db      238  AAGGAACGCGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAGAAATTTGAGAGA 297
Qy      1427  GAAAGCGGAGCTTTTACAGAGCTGTCATTCGATTGGGGTGGAGAGAAAGCGTTTGA 1486
Db      298  GAAAGCGGAGCTTTTACAGAGCTGTCATTCGATTGGGGTGGAGAGAAAGCGTTTGA 357
Qy      1487  GAAGAGCGGAGCGCTGGGTAAAGCAGCAGCTTTTAAACATCAGCAACTTTGACCAAC 1546
Db      358  GAAGAGCGGAGCGCTGGGTAAAGCAGCAGCTTTTAAACATCAGCAACTTTGACCAAC 417
Qy      1547  AACTCAGAAATGTGAAACTTTTCAAGTGCCTTCTCAGGAAGTTCGATCCAGCAATCTT 1606
Db      418  AACTCAGAAATGTGAAACTTTTCAAGTGCCTTCTCAGGAAGTTCGATCCAGCAATCTT 477
Qy      1607  ATAGTCCACTCAGCGCCAGCGCAAAAGACTACACAGTGTGGCTAAATGGGGTGCAGCT 1666
Db      478  ATAGTCCACTCAGCGCCAGCGCAAAAGACTACACAGTGTGGCTAAATGGGGTGCAGCT 537
Qy      1667  TGCACATCAAAACTGACTAAATCTCTCTCGCTCCTCAGTGTGCTGCTGCTGCTGCTG 1726
Db      538  TGCACATCAAAACTGACTAAATCTCTCTCGCTCCTCAGTGTGCTGCTGCTGCTGCTG 597
Qy      1727  ACACATTCATGTGTCTGTAACACACAGTTCCTCAGTGTGCTGCTGCTGCTGCTGCTG 1786
Db      598  ACACATTCATGTGTCTGTAACACACAGTTCCTCAGTGTGCTGCTGCTGCTGCTGCTG 657
Qy      1787  AGTAAACCAAGTGAAGTTCGAAAGAGCAAGCAGGATCAAGTGAAGTGAAGTGAAGTGA 1846
Db      658  AGTAAACCAAGTGAAGTTCGAAAGAGCAAGCAGGATCAAGTGAAGTGAAGTGAAGTGA 717
Qy      1847  CCCAGCTCGCGGAGGGGTGTACAGCGGATGCTCCTCGGCTTCCAGGAGCGCTCACGG 1906
Db      718  CCCAGCTCGCGGAGGGGTGTACAGCGGATGCTCCTCGGCTTCCAGGAGCGCTCACGG 777
Qy      1907  GACCGAGATGACTTACCTTAAATGTGCGGGTGCAGTGTGCTTCCAGATGTGCGCTAGA 1966
Db      778  GACCGAGATGACTTACCTTAAATGTGCGGGTGCAGTGTGCTTCCAGATGTGCGCTAGA 836
Qy      1967  GGAG 1970
Db      837  GGAG 840
```

```
RESULT 13
CX562093
LOCUS      CX562093
DEFINITION  UI-M-FC0-ct2-g-10-0-UI_r1 NIH_BMAP_F00 Mus musculus cDNA clone
ACCESSION  CX562093
VERSION    CX562093.1
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 834)
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
JOURNAL    Sciurognathi; Muroidae; Muridae; Murinae; Mus.
COMMENT    NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
```

Email: cgapbe-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mouseefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

#### FEATURES

Location/Qualifiers  
1..834  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="CS7BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30699705"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_F00"  
/note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dr primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction.  
ligated with Ecor I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is TGAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

#### ORIGIN

Query Match 30.1%; Score 810; DB 9; Length 834;  
Best Local Similarity 99.5%; Pred. No. 1.2e-197;  
Matches 832; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
Qy 346 GGCAAGAGAGAAATTAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGT 405  
Db 1 GGCAAGAGAGAA-TAAATATAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTGT 59  
Qy 406 TCAGCGGAAGAACCTGCTGGCCAGGAGCGTGGAGACACAGAACTTGAAGCTGGGAG 465  
Db 60 TCAGCGGAAGAACCTGCTGGCCAGGAGCGTGGAGACACAGAACTTGAAGCTGGGAG 119  
Qy 466 TGACATGGACCACTGCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTGGAAACGTCAG 525  
Db 120 TGACATGGACCACTGCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTGGAAACGTCAG 179  
Qy 526 GCGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGCTGCAGTGCAGAGACAGAGTTT 585  
Db 180 GCGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGCTGCAGTGCAGAGACAGAGTTT 239  
Qy 586 GCATCAGCTCTGAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAG 645  
Db 240 GCATCAGCTCTGAAGATGAGAAAGATGAGGTACAAAATTTACAAAATATCATAGCCAG 299  
Qy 646 CCGGGCTACTCAGTATATATCATGTGTGAAGAGGAGGCGTGAATATAATAAGCTAAA 705  
Db 300 CCGGGCTACTCAGTATATATCATGTGTGAAGAGGAGGCGTGAATATAATAAGCTAAA 359  
Qy 706 GGAGCGCTGCATCAGCTCGTTATGAACAAGAGAGTAAATAAGTCCATGGATGTTT 765  
Db 360 GGAGCGCTGCATCAGCTCGTTATGAACAAGAGAGTAAATAAGTCCATGGATGTTT 419  
Qy 766 AAATTATGTGGTGCAGCTGATGCAACCGAGGCTCATGGAGGACTGCAAAAACAGAGCC 825  
Db 420 AAATTATGTGGTGCAGCTGATGCAACCGAGGCTCATGGAGGACTGCAAAAACAGAGCC 479

826 CAGGAATGAAGATGAGATGTACAAATTCCTTTGAATGATTATGATACCCAGAGCA 885  
 |||||  
 480 CAGGAATGAAGATGAGATGTACAAATTCCTTTGAATGATTATGATACCCAGAGCA 539  
 |||||  
 886 GATCCTGATGAGAAACGGGAGCTGAAGAGGTCTCCAGCAGATGAAGAGGAGATGAT 945  
 |||||  
 540 GATCCTGATGAGAAACGGGAGCTGAAGAGGTCTCCAGCAGATGAAGAGGAGATGAT 599  
 |||||  
 946 CTCTCTCTGCTCTCTCCTCAGAAAGAACCCAGGGAAGAGCAGAGGACGACAGGCAC 1005  
 |||||  
 600 CTCTCTCTGCTCTCTCCTCAGAAAGAA-CCAGGGAAGAGCAGAGGACGACAGGCAC 658  
 |||||  
 1006 TGTTCCTATCTCCATATAGAGATGACTCTGGGGAAGTGGAGCAGACAGCGTGTGGGG 1065  
 |||||  
 659 TGTTCCTATCTCCATATAGAGATGACTCTGGGGAAGTGGAGCAGACAGCGTGTGGGG 718  
 |||||  
 1066 CTTTTCCTGTCACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAT 1125  
 |||||  
 719 CTTTTCCTGTCACACTGTGAGAGAGCAGCTGACANACAGCATCAGGAAACAGTGGAGAT 778  
 |||||  
 1126 TTTGAAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAGTACACTCAGAGG 1181  
 |||||  
 779 TTTGANAAGTCATGTAGAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGG 834  
 |||||

## RESULT 14

CN539511 804 bp mRNA linear EST 29-APR-2004  
 LOCUS UI-M-HUO-cqu-n-01-0-UI.r1 NIH\_BMAP\_HUO Mus musculus cDNA clone  
 DEFINITION IMAGE:30667224 5', mRNA sequence.

CN539511  
 CN539511.1 GI:46867667

## VERSION

## KEYWORDS

## SOURCE

Mus musculus (house mouse)

## ORGANISM

Mus musculus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 804)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

## FEATURES

## source

1...804  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="CS7BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30667224"  
 /tissue\_type="whole eye"  
 /dev\_stages="newborn (1, 5, 15 days)"  
 /lab\_hosts="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP HUO"  
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned

## RESULT 15

CX562859

LOCUS

DEFINITION

UI-M-IBO-cub-p-21-0-UI.r1 NIH\_BMAP\_IBO Mus musculus cDNA clone

798 bp

mRNA

linear

EST 12-JAN-2005

directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAATACG. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 29.8%; Score 802; DB 8; Length 804;  
 Best Local Similarity 99.8%; Pred. No. 1.3e-195;  
 Matches 802; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 646 CCGGGCTACTCAGTATATCATGTGAAGAGAGGAGCGTGAATATATAAGCTAAA 705  
 |||||  
 1 CCGGGCTACTCAGTATATCATGTGAAGAGAGGAGCGTGAATATATAAGCTAAA 60  
 |||||  
 706 GGAGCGCTCATCAGCTCGTTATGAACAAGAGAGGATAAAAAATAGCCATGGATGTTT 765  
 |||||  
 61 GGAGCGCTCATCAGCTCGTTATGAACAAGAGAGGATAAAAAATAGCCATGGATGTTT 120  
 |||||  
 766 AAATTATGTTGGTTCGAGCTGATGCAACGAGGCTCATGAGGACTGACAAAACAGAGC 825  
 |||||  
 121 AAATTATGTTGGTTCGAGCTGATGCAACGAGGCTCATGAGGACTGACAAAACAGAGC 180  
 |||||  
 826 CAGGAATGAAGATGAGATGTACAAAATTCCTTTGAATGATTATGAGTACCGCCAGAGCA 885  
 |||||  
 181 CAGGAATGAAGATGAGATGTACAAAATTCCTTTGAATGATTATGAGTACCGCCAGAGCA 240  
 |||||  
 886 GATCCTGATGAGAACCGGAGCTGAAGAGGTCTCTCCAGCAGATGAAGAGGAGATGAT 945  
 |||||  
 241 GATCCTGATGAGAACCGGAGCTGAAGAGGTCTCTCCAGCAGATGAAGAGGAGATGAT 300  
 |||||  
 946 CTCTCTCTCTCTCTCTCCTCAGAAAGAGCCAGGGAAGAGCAGAGCAGCAGAGCAC 1005  
 |||||  
 301 CTCTCTCTCTCTCTCTCCTCAGAAAGAGCCAGGGAAGAGCAGAGCAGCAGAGCAC 360  
 |||||  
 1006 TGTTCCTATCTCCGATATAGAAGTACTCTCGGGAACTGAGCAGAGACACGCTGTGGGG 1065  
 |||||  
 361 TGTTCCTATCTCCGATATAGAAGTACTCTCGGGAACTGAGCAGAGACACGCTGTGGGG 420  
 |||||  
 1066 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAT 1125  
 |||||  
 421 CTTTTCCTGTGACACTGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAT 480  
 |||||  
 1126 TTTGAAAAGTCATGTAGAAAACCTCGATACCAAGCTTCGAAAGTACACTCAGAGGCGCT 1185  
 |||||  
 481 TTTGAAAAGTCATGTAGAAAACCTCGATACCAAGCTTCGAAAGTACACTCAGAGGCGCT 540  
 |||||  
 1186 TAATGAGGAGGAGCTCATCTCAGCACAAGACCATGAGCAAGAGACTGAGAAACTGGAGCT 1245  
 |||||  
 541 TAATGAGGAGGAGCTCATCTCAGCACAAGACCATGAGCAAGAGACTGAGAAACTGGAGCT 600  
 |||||  
 1246 GGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCT 1305  
 |||||  
 601 GGAGATTGAGCGGTGTTAAAGAGATGATCAAGGCTCAGCAGAGCTCTTACAGCAGCAGCT 660  
 |||||  
 1306 GGCCACCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1365  
 |||||  
 661 GGCCACCACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
 |||||  
 1366 AAAGGAAACCGCTTTAAAGAGAGTGGACCCCTTTTAAAGAGACAAAAAAGAAATTTTGAGAG 1425  
 |||||  
 721 AAAGGAAACCGCTTTAAAGAGAGTGGACCCCTTTTAAAGAGACAAAAAAGAAATTTTGAGAG 780  
 |||||  
 1426 AGAAAGCGGAAGCTTTACAGAAAGC 1449  
 |||||  
 781 AGAAAGCGGAAGCTTTACAGAAAGC 804  
 |||||

IMAGE:6807142 5', mRNA sequence.  
 ACCESSION CX562859  
 VERSION CX562859.1 GI:57589888  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridea; Muridae; Murinae; Mus.  
 1 (bases 1 to 798)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousef1.html>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.  
 Location/Qualifiers  
 1..798  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6807142"  
 /tissue\_type="whole eye"  
 /dev\_stages="newborn (1,5,15 days)"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP IB0"  
 /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAATACG. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 29.6%; Score 798; DB 9; Length 798;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-194;  
 Matches 798; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 150 CAGAAACAAGATGTCCTCCGTCAGTTGTACTCCAGCAAGTTCTGTGCTCTTCAGTAC 209  
 1 CAGAAACAAGATGTCCTCCGTCAGTTGTACTCCAGCAAGTTCTGTGCTCTTCAGTAC 60  
 210 CTTTATCCAAACAGTCATGCTGTTTTCGGTGTCTTCTGCACAGGAGAGAACATTGAAC 269  
 61 CTTTATCCAAACAGTCATGCTGTTTTCGGTGTCTTCTGCACAGGAGAGAACATTGAAC 120  
 270 AAAGTATTTCTTATCTTTGATCAGGAGCTGACACCTTCGGGTTCCTTCCTTGTATGAAG 329  
 121 AAAGTATTTCTTATCTTTGATCAGGAGCTGACACCTTCGGGTTCCTTCCTTGTATGAAG 180  
 330 AATCCAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTTCTGAATGTATGA 389  
 181 AATCCAAAGTAAAGAGCAAGAGAGAAATTAATATAGTCGCTGTTCTGAATGTATGA 240  
 390 ACGAGCTGCTGCTCTTCAGCGGAAGAACCTGCTGGCCCGAGGAGCGTGGAGACACAGA 449

Db 241 ACGAGCTGCTGCTGCTTCAGCGGAAGAACCTGCTGGCCCGAGAGAGCGTGGAGACACAGA 300  
 QY 450 ACTTGAAGCTGGGCAGTGCATGGACCACTGCGAGAGCTGCTACGCCAAACTTTAAGGAGC 509  
 Db 301 ACTTGAAGCTGGGCAGTGCATGGACCACTGCGAGAGCTGCTACGCCAAACTTTAAGGAGC 360  
 QY 510 AGTTGGAAACCTCCAGGGCGGAGATGATCGGGCTTCAAGAGAGACAGGCGAGCTGCAGT 569  
 Db 361 AGTTGGAAACCTCCAGGGCGGAGATGATCGGGCTTCAAGAGAGACAGGCGAGCTGCAGT 420  
 QY 570 GCAAGAACAGGAGTTTGGTCATCAGCTCCCTGGAAGAATGAGAAAGATGAGTACAAAAATTAC 629  
 Db 421 GCAAGAACAGGAGTTTGGTCATCAGCTCCCTGGAAGAATGAGAAAGATGAGTACAAAAATTAC 480  
 QY 630 AAAATATCATAGCCAGCGGGCTACTCAGTATATAATCATGATGTCGAAGAGGAGGAGCGTG 689  
 Db 481 AAAATATCATAGCCAGCGGGCTACTCAGTATATAATCATGATGTCGAAGAGGAGGAGCGTG 540  
 QY 690 AATATATAAGCTAAAGGAGCGCTGCATCAGCTCGTTTATGAACAAAGAGGATAAAAACA 749  
 Db 541 AATATAATAAGCTAAAGGAGCGCTGCATCAGCTCGTTTATGAACAAAGAGGATAAAAACA 600  
 QY 750 TAGCCATGGATGTTTAAATTTATGCTGGCTCGAGCTGATGGCAACGAGGCTCATCGAGGA 809  
 Db 601 TAGCCATGGATGTTTAAATTTATGCTGGCTCGAGCTGATGGCAACGAGGCTCATCGAGGA 660  
 QY 810 CTGACAAAACAAGAGCCAGGAATGAAGATGAGATGTTACAAAATTTCTGTTGAATGATTATG 869  
 Db 661 CTGACAAAACAAGAGCCAGGAATGAAGATGAGATGTTACAAAATTTCTGTTGAATGATTATG 720  
 QY 870 AGTACCCCGCAGAACAGATCCTGATGGAGAACCGGAGCTGGAAGAGTCTTCCAGCAGAGA 929  
 Db 721 AGTACCCCGCAGAACAGATCCTGATGGAGAACCGGAGCTGGAAGAGTCTTCCAGCAGAGA 780  
 QY 930 TGAAGAAAGGAGATCATCT 947  
 Db 781 TGAAGAAAGGAGATCATCT 798

Search completed: June 14, 2006, 04:03:37  
 Job time : 11424 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 13, 2006, 23:24:10 ; Search time 7913 Seconds  
(without alignments)  
6519.093 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVPVLTENKNLS.....CYSGCSSAFSAHGDRDDLP 615

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 48236798 seqs, 27959665780 residues

Total number of hits satisfying chosen parameters: 96473596

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlh  
-Q=/abses/ABSWEB\_spool/US10644084/runat\_12062006\_150418\_1399/app\_query.fasta\_1  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISP=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abses08  
-USER=US10644084 @CGN\_1\_1\_6323 @runat\_12062006\_150418\_1399 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est3:\*  
3: gb\_est4:\*  
4: gb\_est5:\*  
5: gb\_est6:\*  
6: gb\_hcc:\*  
7: gb\_est2:\*  
8: gb\_est7:\*  
9: gb\_est8:\*  
10: gb\_est9:\*  
11: gb\_gss1:\*  
12: gb\_gss2:\*  
13: gb\_gss3:\*  
14: gb\_gss4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3150.5	99.5	1845	AY412493	AY412493 Mus muscu
2	3149.5	99.5	3185	AK043865	AK043865 Mus muscu
3	3149.5	99.5	3413	AK031356	AK031356 Mus muscu
4	3140.5	99.2	2245	AK077112	AK077112 Mus muscu

5	3138	99.1	2371	6	AK049080
6	2786.5	88.0	1845	14	AY412491
7	2786.5	88.0	2111	6	CR613292
8	2786.5	88.0	3154	6	BC035580
9	2786.5	88.0	5658	6	BSM801317
10	2747.5	86.8	1845	14	AY412492
11	1482	46.8	878	9	CX565842
12	1416.5	44.8	903	10	DV893190
13	1364	43.1	804	8	CN539511
14	1356	42.8	834	9	CX562093
15	1349.5	42.6	840	5	CF949831
16	1348	42.6	798	9	CX562859
17	1323.5	41.8	933	4	CB205832
18	1314	41.5	822	8	CV563162
19	1300.5	41.1	891	9	DN286111
20	1291	40.8	771	3	BQ179212
21	1287	40.7	860	10	DT831879
22	1263	39.9	770	5	CF729394
23	1251	39.5	760	8	CN528866
24	1235	39.0	746	8	CN528831
25	1233.5	39.0	845	3	BQ769157
26	1221	38.6	723	5	CF737957
27	1216.5	38.4	1005	1	AL582610
28	1216	38.4	828	10	DV803752
29	1208	38.2	712	5	CF732776
30	1199.5	37.9	800	9	DN877534
31	1196.5	37.8	833	8	CV117234
32	1192	37.7	825	1	AU133150
33	1184	37.4	954	1	AL561392
34	1168	36.9	736	8	CX226608
35	1162	36.7	930	4	BX372332
36	1153	36.4	771	5	CF737129
37	1149	36.3	687	9	CX568194
38	1145	36.2	679	8	CN532570
39	1141	36.1	683	8	CO424815
40	1140.5	36.0	733	9	DN864383
41	1076.5	34.0	805	2	BG242849
42	1075.5	34.0	638	8	CN720771
43	1069	33.8	751	5	CJ465026
44	1062	33.6	640	5	CF737822
45	1056	33.4	748	8	CO555994

ALIGNMENTS

RESULT 1	AY412493	1845 bp	DNA	linear	GSS 12-DEC-2003
LOCUS	AY412493	Mus musculus HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,			
DEFINITION	AY412493	Genomic survey sequence.			
ACCESSION	AY412493	GI:39768458			
VERSION	AY412493.1	GI:39768458			
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 1845)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios				
JOURNAL	Science 302 (5652), 1960-1963 (2003)				
PUBMED	14671302				
REFERENCE	2 (bases 1 to 1845)				
AUTHORS	Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarawal, A., Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J., Adams, M.D. and Cargill, M.				
TITLE	Direct Submission				

JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES  
source Location/Qualifiers  
1..1845  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
gene <1..>1845  
/locus\_tag="HCM4575"  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,266-304 Length: 1845  
Score: 3150.50 Matches: 614  
Percent Similarity: 99.8% Conservative: 0  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 99.5% Indels: 1  
DB: 14 Gaps: 1  
US-10-644-084-2 (1-615) x AY412493 (1-1845)  
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
DB 1 ATGGGAGATTGGATGACTGCTGACAGATCCAGTTCTGTGTACAGAAAACAAAAATCTCTCT 60  
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
DB 61 CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTTGTACTCCAGCAAGTTCTGTGC 120  
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
DB 121 TCTTCAGTACCTTTATCCAAAACGTCATGTGTGTTTCGGGTGTTCTTCGACAGAGAG 180  
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
DB 181 AACATTGAACAAAGTATTTCCTATCTTGATCAGGAGCTGACCACCTTCGGGTTCCTTCC 240  
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
DB 241 TTGTATGAAGAAATCCAAAAGTAAAGAGGCAAGAGAGAAATTAATATATAGTCGCTGTTCTG 300  
QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
DB 301 AACTGTATGAACGCTGCTGCTTTCAGCGGAAGAACCTGCTGCTCCAGAGAGACGCTG 360  
QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
DB 361 GAGACACAGAACTTGAAGCTGGGCAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA 420  
QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
DB 421 CTTAAGGAGCAGTTGGAAACGTCACGCGGGAGATGATCGGGCTTCAAGAGAGACACAGG 480  
QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
DB 481 CAGCTGCGATGCAAGAACAGGAGTTTGCATCAGCTCTCTGAAGAAATGAGAAAGATGAGGTA 540  
QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
DB 541 CAAAAAATTACAAAATATCATAGCCAGCGGGCTACTCAGTATATATCATGTGTGAAGAGG 600  
QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
DB 601 AAGGAGCGGTGAATATAAATAAGCTAAAGGAGCGCCCTGCATCAGCTCGTTATGAACAAGAG 660  
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
DB 661 GATAAATAACATAGCCATGGAATGTTTAAATATATGTGGTCTGAGCTGTATGGCNAACAGGC 720  
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260

DB 721 TCATGGAGGAGCTGACAAAACAGAACCCAGGAATGAAGATGAGATGTCACAAAATTCGTGTG 780  
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280  
DB 781 AATGATTATGATACCGCCAGAACAGATCTGTATGGGAACCGCGAGCTGAAGAGGTC 840  
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
DB 841 CTCAGCAGATGAAGAGGAGATGATCTCTCTCTCTCTCTCTCAGAGAGAGAGCCAGG 900  
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
DB 901 GAAAGAGCAGAGGACGCGCACAGGCACCT--GCTATCTCCGATATAGAAGATGACTCTGGG 957  
QY 321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
DB 958 GAACGTGACAGACAGACGCGTGTGGGCGCTTTCCTGTGTGACCTGTGTGAGAGAGCAGCTGACA 1017  
QY 341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
DB 1018 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAAATCGATAACCAA 1077  
QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
DB 1078 GCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGGACGTCATCTCACGACACAGCCAT 1137  
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
DB 1138 GAGCAAGAGACTGAGAACTGGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1197  
QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
DB 1198 CAGCAGCAGCTCTTACAGCAGCAGCTGGCCACCACGCTGTGTATGATGACACACCTCCTG 1257  
QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPhe 440  
DB 1258 TTGGCAGACTGTTACTTGTCTGGAAGAAAGAAAGCGCTTAAAGAGAGAGTGGACCCCTTTT 1317  
QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460  
DB 1318 AAAGAGCAAAAAAGAAATTTTGAAGAGAAAGGCGAGCTTTTACAGAGCTGCCAATCGA 1377  
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnPhe 480  
DB 1378 TTGGGTTTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCGACCTGGGTAAAGCAGCAGTTT 1437  
QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500  
DB 1438 TTAACACATGACGAACCTTTGACCCACGAACTCAGAAAAATGTGAAACTTTTTCAGTGCCTTC 1497  
QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
DB 1498 TCAGAAAGTTCTGATCCAGACAACTTATATAGTCCACTCAGCGCCACGCCAAAAGAGCTA 1557  
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540  
DB 1558 CACAGTGTGGCTAATGGGTGCCAGCTTGACATCAAAACTGACTAAATCTCTTCTGCTGCC 1617  
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560  
DB 1618 TCACCTTCTACTTCTCAGACTTTTCGCGCAGACACATTCATGTGTCTGTGAACACAGATTCCATC 1677  
QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580  
DB 1678 AGTGTGCTGAATATAACTCTCTGAAGAAAGTAAACCAAGTGAAGTTTGAAGAGAAAGCAGG 1737  
QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600  
DB 1738 GATCAGAAGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGGAGGGGTGCTACAGCGGATGC 1797  
QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
DB 1798 TCCTCGGCTTTTCAGAGCGCTTCACGGGACCGAGATGATGACTTACCT 1842

## RESULT 2

## AK043865

## LOCUS

## DEFINITION

AK043865 3185 bp mRNA linear HTC 02-SBP-2005  
Mus musculus 10 days neonate cortex cDNA, RIKEN full-length  
enriched library, clone:A830043F14 product:HYPOHETICAL 71.0 KDA  
PROTEIN homolog [Mus musculus], full insert sequence.

## ACCESSION

## AK043865

## VERSION

## AK043865.1

## KEYWORDS

## HTC; CAP trapper.

## SOURCE

## Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## 1

## Carninci, P. and Hayashizaki, Y.

## High-efficiency full-length cDNA cloning

## Meth. Enzymol. 303, 19-44 (1999)

## PUBMED

## 10349636

## REFERENCE

## 2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes

## Genome Res. 10 (10), 1617-1630 (2000)

## PUBMED

## 11042159

## REFERENCE

## 3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer

## Genome Res. 10 (11), 1757-1771 (2000)

## PUBMED

## 11076861

## REFERENCE

## 4

The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.

## Functional annotation of a full-length mouse cDNA collection

## Nature 409, 685-690 (2001)

## PUBMED

## 11076861

## REFERENCE

## 5

The FANTOM Consortium, the RIKEN Genome Exploration Research Group  
Phase I and II Team.Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs

## Nature 420, 563-573 (2002)

## PUBMED

## 11076861

## REFERENCE

## 6

RIKEN Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium.

## Antisense Transcription in the Mammalian Transcriptome

## Science 309, 1564-1566 (2005)

## PUBMED

## 11076861

## REFERENCE

## 7

The FANTOM Consortium, Riken Genome Exploration Research Group and  
Genome Science Group (Genome Network Project Core Group).

## The Transcriptional Landscape of the Mammalian Genome

## Science 309, 1559-1563 (2005)

## PUBMED

## 11076861

## REFERENCE

## 8

## (bases 1 to 3185)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayashida, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koyas, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

## TITLE

## JOURNAL

Direct Submission

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,  
Kanagawa, 230-0045, Japan [E-mail: genome-res@gsc.riken.jp,  
URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222,  
Fax: 81-45-503-9216]

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN,  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.jp/  
URL: http://fantom.gsc.riken.jp/.

## FEATURES

## Source

1..3185  
/location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:A830043F14"  
/db\_xref="taxon:10090"  
/clone="A830043F14"  
/tissue\_type="cortex"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"  
389..2233  
/note="unnamed protein product; HYPOHETICAL 71.0 KDA  
PROTEIN homolog [Mus musculus] (SPT:RAH21749, evidence:  
FASTV, 99.8%ID, 100%length, match=1842)  
putative"  
/codon\_start=1  
/protein\_id="BAC31684.1"  
/db\_xref="GI:26335972"  
/translation="MGDMVTVDPLCTENKLSQYTSKSPSLYSQOVLCSVP  
LSKNVHGVFGCTGENIEQISLYLDQELTTFGFSLEYESKSKARLENIVAVLNC  
MNEELLQRKLLAESVETQNLKGDMDHLSQYAKLKEQLETSSRMELGLQDRD  
LOCKNRSLLHQLNKDEQVQLNIIASRATQYNNHVKRERENKLERLHQLVMN  
KKXNIAQMDLVNIVGRADGKSGMRTDKTEARNEDMYKLLNDYERQKQILMENAE  
LKVLQOAMKMSMISLLSPQKKPRERAEDGTCTVAISDIEDDSGLSRDSVGLSCDT  
VREQLTNSIRQWRLKSHVEKLDNQAKVHSEGLNEEDVISRQHEQTEKLELEIE  
RCKMTAKQQLQALATTCDDTTSLLRDCLYLLBEKKEKLEWTLFQKQKNFERER  
RSFTAAIRLGLERKAFEEERASWYKQFLNNMTNFHQNSENVKLFSAFGSSDPDL  
IVHSRPROKHLHSHVANGVPACTSKLTKSLPASPTSDFRQTHSCVSEHSSISVLTTP  
EESKPEVAERSTDOKWSVQSRPSREGCYSCSSAFRSAGHGRDRLP"

## CDS

polyA\_signal  
polyA\_site  
/note="putative"  
3185  
/note="putative"

## ORIGIN

Alignment Scores:  
Pred. No.: 9,72e-304 Length: 3185  
Score: 3185.50 Matches: 614  
Percent Similarity: 99.8% Conservative: 0  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 99.5% Indels: 1  
DB: 6 Gaps: 1

US-10-644-084-2 (1-615) x AK043865 (1-3185)

Qy 1 MetGlyAspTTPMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
Db 389 ATGGGAGATTGGATGACATGTCACAGATCCAGTCTGTGTACAGAAAACAAATCTCTCT 448  
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Db 449 CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGTTTGTACTCCAGCAGTCTCTGTGC 508  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 509 TCTTCAGTACCTTTTATCAAAAACGTCATGCTGTTTTCGGGTCTTCTTGCACAGAGAG 568

Qy 61 AenilleGluGlnSerIleSerTyrLeuAepGlnGluLeuThrThrPheGlyPheProSer 80  
Db AACATTGAACAAGTATTTCCTATTCTTGNATCAGGAGCTGACCACTTCGGGTTCCTTCC 628  
Qy 81 LeutyfGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Db TTGTATGAAGAATCCAAAGTAAAGAGGCAAGAGAGAATTAATATATAGTCGCTCTCTG 688  
Qy 101 AenCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerVal 120  
Db AACTGTATGACAGCTGCTCGTGTCTCAGCGGAAGAACCTGCTGCCCGAGAGAGCGTG 748  
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db GAGACACAGAACTTGAAGCTGGCGAGTGNATGACCACTGCAGAGCTGTACGCCAAA 808  
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Db CTTAAGGAGCAGCTTGAAACGTCACAGCGGAGATGATCGGCTTCAAGAGAGAGACAGG 868  
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180  
Db CAGCTGCGAGTGCAGAAACAGGAGTTTGCATCAGCTCCCTGAAGAATGAAGAAGATGAGTA 928  
Qy 181 GlnLysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
Db CAANAATATACAAATATATACAGCCAGCGGCTACTCAGTATATATCATGATGTGAAGAG 988  
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Db AAGGAGCGTGAATATAATAAGCTAAAGAGCGCTGCATCAGCTCGTATGAACAAGAG 1048  
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db GATAAAACATAGCATGATGATGTTTTAAATTATGTGGTTCGAGCTGATGGCAACAGGCG 1108  
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Db TCATCGAGGAGCTGACAAACAGAGCCAGGAGTGAAGATGAGATGTACAAATATCTGTG 1168  
Qy 261 AenAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280  
Db AATGATTATGATACCGCAGAGCAGATCCTGTATGAGAGAACCGCGAGCTGAAGAAGTGC 1228  
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
Db CTCACGACAGTGAAGAAGGAGATGATCTCTCTGTCTCTCAGAAAGAAAGCCACAGG 1288  
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
Db GAAAGGAGGAGGAGGAGGAGGAGCTGTGCTATCTCCGATATAGAGATGACTCTGGG 1348  
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
Db GAACTTCAGCAGACAGACGCGTGTGGGCGCTTTCCTGTGACACTGTGAGAGAGCAGCTGACA 1408  
Qy 341 AenSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
Db AACACATCAGGAACACAGTGGAGAATTTTGAAGAGTCATGTAGAAAACTCGATAACCA 1468  
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
Db GCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGGAGCGTCATCTCAGCAACAGCCAT 1528  
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluLysGluArgCysLysGluMetIleLysAla 400  
Db GAGCAAGAGACTGAGAAACTGAGAGCTGGAGATTGAGCGGTGTAAAGAGATGATCAAGGCT 1588  
Qy 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
Db CAGCAGCAGCTCTTA---CAGCAGCTGCCACCACCTGTGTGATGATGACACCACTCTACTG 1645

Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440  
Db TTGCGAGACTGTACTTGTCTGNAAGAAAGAAAGCGCTTAAAGAAAGAGTGGACCTTTT 1705  
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460  
Db AAAGAGCAAAAAAGAAATTTTGAGAGAGAAAGGCGAAGCTTTACAGAAAGCTGCCATTGA 1765  
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnPhe 480  
Db TTGGGTTTGGAGAGAAAGCGCTTTGAAGAGAGCGAGCGAGCTGGTGAAGCAGCAGTTT 1825  
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500  
Db TTAAACATGACGAACTTTTGACCACACAGAACTCAGAAAAATGTGAAACTTTTCAGTGCCTTC 1885  
Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
Db TCAGAAAGTTCTGATCCAGACAATCTTATAGTCCACTCACGCCACCGCAAAAGAGCTTA 1945  
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540  
Db CACAGTGTGGCTAATGGGTGCCAGCTTGCACATCAAAACTGACTAAATCTCTTCTGCC 2005  
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560  
Db TCACCTTCTACTCTCAGACTTTTCGCCAGACACATTCATGTGTCTGAAACACAGATTCATC 2065  
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580  
Db AGTGTGCTGAATATAACTCTCTGAAGAAAGTAAACCAAGTGAAGTTGCAAGAGAAAGCAG 2125  
Qy 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600  
Db GATCAGAAAGTCGACGCTGTCAGTCGAGGCCAGCTCGCGGAGGGGTGTACAGCGGATGC 2185  
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
Db TCCTCGGCTTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 2230  
RESULT 3  
AK031356 3413 bp mRNA linear HTC 02-SEP-2005  
LOCUS Mus musculus 13 days embryo male testis cDNA, RIKEN full-length  
DEFINITION enriched library, clone:603041112 product:hypothetical protein,  
full insert sequence.  
ACCESSION AK031356  
VERSION AK031356.1 GI:26327238  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
REFERENCE  
1 Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2  
3 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3  
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,  
Konno,H., Akiyama,J., Nishi,K., Kitsuina,T., Tashiro,H., Itoh,M.,  
Sumi,N., Ishii,Y., Nakamura,S., Hazama,N., Nishine,T., Harada,A.,  
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kaehiwagi,K.,  
Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Warahiki,M.,  
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,



Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861

4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5  
The FANTOM Consortium, the RIKEN Genome Exploration Research Group  
Phase I and II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

6  
RIKEN Genome Exploration Research Group, Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium.  
Antisense Transcription in the Mammalian Transcriptome  
Science 309, 1564-1566 (2005)

7  
The FANTOM Consortium, Riken Genome Exploration Research Group and  
Genome Science Group (Genome Network Project Core Group).  
The Transcriptional Landscape of the Mammalian Genome  
Science 309, 1559-1563 (2005)

8 (bases 1 to 3413)  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,  
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,  
Hayashida,K., Hayato,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,  
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,  
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,  
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,  
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
Takada,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa, 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,  
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL:http://genome.gsc.riken.jp/  
URL:http://phantom.gsc.riken.jp/  
Location/Qualifiers  
1. .3413  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:603041112"  
/db\_xref="taxon:10090"  
/clone="603041112"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="13 days embryo"  
264. .2108  
/note="unnamed protein product; hypothetical protein  
(evidence: ProCrest)  
putative"  
/codon\_start=1  
/protein\_id="BAC27363.1"

CDS

/db\_xref="GI:26327239"  
/translation="MGDMWTVTDPVLCTENKNLSQVTSFTKMPSPSLYSQQVLCSSVP  
LSKNHGVGFCTGSEIEQISYSDQELTTFGFSLSYEESKSEAKRELNTVAVLNC  
MNEELVLRKNLLKQESVETQNLKLGSDMDHQLQSCVAKLKEQLETSSRRRMIGLQEDR  
OLOCKNRSLLHLLKNEKDEVOKLQNI IASRATQYNHDKVRKREYNKLRHQLVNR  
KKKNIAMDVLNVGRADGKSGWETDTEARNEDEMYKILLNDVEYRKQILMENAE  
LKKVLQMKKEMIRLSKQKKPRERAEDGTGTVAISDIEDSGLSRUSRWGLSCDT  
VREQJTSIRKQWIRILKSHVEKLDNQASKVSEHSEGNEDVISRQDHEQTEKLBLEIE  
RCKEMIKAQQLQLLATDDDTTSLLRDCLLLEKRLKEWTLFKQKQKFRER  
RSFTEAAIRLGLERKAFBEERASWVKQFPLNMTNFDHQNSENKLPFSAFSGSDPNL  
I VHSRPROKHLHSVANGVPACTSKLTSLPASPSDPROTHSCVSEHSS ISVLNITP  
EESKFSVARESTDQKWSVQSRPSSREGCYSCSSAFRSRSHGDRDLP"  
3394. .3399  
/note="putative"  
3413  
/note="putative"

polyA\_signal  
polyA\_site

ORIGIN

Alignment Scores:  
Pred. No.: 1.08e-303 Length: 3413  
Score: 3149.50 Matches: 614  
Percent Similarity: 99.8% Conservativeness: 0  
Best Local Similarity: 99.8% Mismatches: 0  
Query Match: 99.5% Indels: 1  
DB: 6 Gaps: 1

US-10-644-084-2 (1-615) x AK031356 (1-3413)

Qy 1 MetGlyAspTyrMetThrValThrAspProValLeuLeuGluLeuGluLeuGluLeuSer 20  
Db 264 ATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTACAGAAAACAAAATCTCTCT 323  
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuGluCys 40  
Db 324 CAATATATCTCAGAAACAAAGATGCTCGTCCAGTTGTACTCCAGCAGAGTTCTGTGTC 383  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 384 TCTTCAGTACCTTTATCCAAAACGTCATGCTGTTGTTTGGTGTCTTCTGTCACAGAGAG 443  
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
Db 444 AACATTGAACAAAGTATTTCTTATCTTGTATCAGGAGCTGACCACCTTCGGGTTCCTTCC 503  
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Db 504 TTGTATGAAGAATCCAAAAGTAAAGAGCAAGAGAGAGAAATTAATATAGTCTGCTGTTCTG 563  
Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
Db 564 AACTGTATGACGAGCTGCTGCTGCTTCCAGCGGAAGAACCTCTCGCCAGGAGAGCGTG 623  
Qy 121 GluThrGlnAsnLeuLysGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db 624 GAGACACAGAACTTGAAGCTGGCGAGTGACATGGACCATCTCAGAGCTGCTACGCCAAA 683  
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Db 684 CTTAAGGAGCAGTTGGAAACGTCACGCGGAGAGATGATCGGGCTTCAAGAGAGACAGG 743  
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
Db 744 CAGCTGCAGTGCAGAACACAGGAGTTTGATCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 803  
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
Db 804 CAAAATTTACAAAATATCATGCCAGCCGGCTACTCAGTATATATATATATATATATATAT 863  
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Db 864 AAGGAGCGTGAATATAATAAGCTAAAGGAGCGCTGCATCAGTCTGTTATGAACAAGAAG 923  
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240

[illegible]

Db	2001	GATCAGAAGTGGAGCTGTCAGTCGAGCCCGAGCTCCGGGAGGGGTCTACAGCGGATGC	2061
Qy	601	Seq:SerAlaPheArgSerAlaHisGlyAspArgAspLeuPro	615
Db	2061	TCTCTCGCCCTTCAGGAGCGCTACGGGGACCGAGATGACTTACCT	2105
RESULT 4	AK077112	2245 bp	linear
LOCUS	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933406N16 product:HYPOTHETICAL 71.0 KDA PROTEIN homolog [Mus musculus], full insert sequence.		
ACCESSION	AK077112.1	GI:26097206	
VERSION	HTC; CAP trapper.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.		
REFERENCE	1	Carninci, P. and Hayashizaki, Y.	
AUTHORS	High-efficiency full-length cDNA cloning		
TITLE	Meth. Enzymol. 303, 19-44 (1999)		
JOURNAL	10349636		
PUBLISHED	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.	
REFERENCE	Normalisation and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
TITLE	Genome Res. 10 (10), 1617-1630 (2000)		
JOURNAL	11042159		
PUBLISHED	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohata, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.	
AUTHORS	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
TITLE	Genome Res. 10 (11), 1757-1771 (2000)		
JOURNAL	11076861		
PUBLISHED	4	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.	
REFERENCE	Functional annotation of a full-length mouse cDNA collection		
AUTHORS	Nature 409, 685-690 (2001)		
TITLE	5	The FANTOM Consortium, the RIKEN Genome Exploration Research Group Phase I and II Team.	
JOURNAL	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
PUBLISHED	Nature 420, 563-573 (2002)		
REFERENCE	6	RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Core Team) and the FANTOM Consortium.	
AUTHORS	Antisense transcription in the Mammalian Transcriptome Science 309, 1564-1566 (2005)		
TITLE	7	The FANTOM Consortium, Riken Genome Exploration Research Group and the Genome Science Group (Genome Network Project Core Group).	
JOURNAL	The Transcriptional Landscape of the Mammalian Genome Science 309, 1559-1563 (2005)		
PUBLISHED	8 (bases 1 to 2245)		
REFERENCE	9	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, S., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,	

Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akaira, S., Takeda, Y., Tanaka, T., Tonaru, A., Toya, T., Yaeunishi, A., Muramatsu, M. and Hayashizaki, Y.

# **TITLE**

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan (E-mail: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT**

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

Location/Qualifiers

## **FEATURES**

Source

1. .2245

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:4933406N16"

/db\_xref="taxon:10090"

/clone="4933406N16"

/sex="male"

/tissue\_type="testis"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="adult"

1. .2245

/note="HYPOTHEtical 71.0 KDA PROTEIN homolog (Mus musculus) (SPTR|AAH21749, evidence: FASTY, 99.7%ID, 100%length, match=1842)

putative"

misc\_feature

## **ORIGIN**

Alignment Scores:

Pred. No.: 4, 47e-303 Length: 2245

Score: 3140.50 Matches: 613

Percent Similarity: 99.7% Conservative: 0

Best Local Similarity: 99.7% Mismatches: 1

Query Match: 99.2% Indels: 1

DB: 6 Gaps: 1

US-10-644-084-2 (1-615) x AK077112 (1-2245)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
 DB ATGGGAGATTGGATGACTGTGACAGATCCAGTTCGTGTACAGAAAACAAAATCTCTCT 384  
 QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys 40  
 DB CAATATACCTCAGAAACAAAGATGTCTCCGTCCAGTTTGTACTCCAGCAAGTTCTGTGC 444  
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
 DB TCTTCAGTACCTTATCCAAAACAGTCATGGTGTTCGGTGTCTTTCGACAGAGAG 504  
 QY 61 AsnileGluGlnSerileSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
 DB AACATTGAACAAAGATTTCTCTATCTTGATCAGGAGCTGACCACCTTCGGGTTTCTTCC 564  
 QY 81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValValLeu 100  
 DB TTGTATGAAGAATCCAAAAGTAAAGAGGCAAGAGAGAGATTAATATATAGTCGTCTCTG 624  
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
 DB AACTGTATGAACAGCTGCTGTCTTTCAGCGGAAGAACCTCTGCGCCAGGAGCGGTG 684

QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
 DB GAGACACAGAACTTGAAGCTGGCAGTGACATGGACCACCTGCGAGAGCTGCTACGCCAAA 744  
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 DB CTTAAGGAGCAGTTGGAAACGTCACGCGGAGATGATCGGGCTTCAAGAGAGAGACAGG 804  
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
 DB CAGCTGCAGTGCACAAAGACAGAGGTTTGCATCAGCTCTCGAGNATCAGAAAGATGAGGTA 864  
 QY 181 GlnLysLeuGlnAsnIleIleIleSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
 DB CAAAAATTACAAAATATCATAGCCGCGGTCTACTCAGTATATAATCATGTGTGAAGAGG 924  
 QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
 DB AAGGAGCGTGAATATAATAAGCTAAAGAGCGCCCTGCATCAGCTCGTTATGAACAAGAAG 984  
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
 DB GATAAAAACATAGCCATGATGTTTAAATTATGGGTGCGAGCTGATGGCAACAGAGGC 1044  
 QY 241 SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
 DB TCATCGAGGAGACTGACAAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATCTCTGTTG 1104  
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
 DB AATGATTATGAGTACCGCCAGAGCAGATCCTGATGGAGAACCGCGAGCTGGAAGAGGTC 1164  
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
 DB CTCTAGCAGATGAGAGAGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1224  
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320  
 DB GAAAGAGCAGAGGACGGCAGCAGCTGTTGCTATCTCCGATATAGAAGATGACTCTGGG 1284  
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
 DB GAACTGAGCAGACAGACGCGTGTGGGGCTTTCTCTGTGACACTGTGAGAGAGCAGCTGACA 1344  
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
 DB AACACATCAGGAAACAGTGGAGAAATTTGAAAAGTCTGTAGAAAACTCGAATAACCAA 1404  
 QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
 DB GCTTCGAAGGTACACTCAGAGGGCTTAAATGAGGAGGAGCTCATCTCACGACAAGACCAT 1464  
 QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
 DB GAGCAAGAGACTGAGAAAACGAGAGCTGAGATTTGAGCGGTGTAAGAGAGATGATCAAGGCT 1524  
 QY 401 GlnGlnGlnLeuGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu 420  
 DB CAGCAGCAGCTCTTAA---CAGCAGCTGGCCACCGCTGTGATGATGACACACCTCCTCTG 1581  
 QY 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLysGluGluTrpThrLeuPhe 440  
 DB TTTCGAGACTGTTACTCTCTCGAAGAAAGAAACGCTTAAAGAGAGAGTGGACCTCTTTT 1641  
 QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaIleArg 460  
 DB AAAGCAAAAAAAGAAATTTTGAAGAGAAAGGCGAAGCTTTTACAGAAGCTGCCATTCGA 1701  
 QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnPhe 480  
 DB TTGGGTTTGGAGAGAAAGCGCTTTTGAAGAGAGCGCAGCTGCTGGGTAAAGCAGCAGTTT 1761

QY	481	LeuAanMetThrAsnPheAspHisGlnAenSerGluAenValLysLeuPheSerAlaPhe	500
DB	1762	TTAAACATGACGAACTTTGACCACAGAACTTCAGAAATGTGAAACTTTTCAGTGCCTTC	1821
QY	501	SerGlySerSerAspProAspHisLeuValHisSerArgProArgGlnLysLysLeu	520
DB	1822	TCAGGAAGTCTTGATCCGACAACTCTTAGTCCACTCAGCGCCACGCGCAAGAAAGCTA	1881
QY	521	HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla	540
DB	1882	CACATGTGGCTTAATGGGTGCCAGCTTGCACATCAAACTGACTAAATCTTCTCCTGCC	1941
QY	541	SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle	560
DB	1942	TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTCTGAACACAGATTCATC	2001
QY	561	SerValLeuAsnIleThrProGluSerLysProSerGluValAlaAArgGluSerThr	580
DB	2002	AGTGTGCTGAATATAACTCTCTGAAGAAAGTAAACCAAGTGAGGTTCGAAGAGAAACGACG	2061
QY	581	AepGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTySerGlyCys	600
DB	2062	GATCAGAAGTGGAGCGTGCAGTCGAGGCCACCTCGCGGAGGGGTGTACAGCGGATGC	2121
QY	601	SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro	615
DB	2122	TTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT	2166
RESULT 5			
AK049080			
LOCUS			
DEFINITION	AK049080	2371 bp mRNA linear HTC 02-SEP-2005	
		Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length	
		enriched library, clone:C230097E14 product:hypothetical protein,	
		full insert sequence.	
ACCESSION	AK049080		
VERSION	AK049080.1	GI:26339745	
KEYWORDS		HTC; CAP trapper.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	
		Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			

TITLE	Phase I and II Team.
JOURNAL	Analysis of the mouse transcriptome based on functional annotation
REFERENCE	of 60,770 full-length cDNAs
AUTHORS	Nature 420, 563-573 (2002)
TITLE	RIKEN Genome Exploration Research Group, Genome Science Group
JOURNAL	(Genome Network Core Team) and the FANTOM Consortium.
REFERENCE	Antisense Transcription in the Mammalian Transcriptome
AUTHORS	Science 309, 1564-1566 (2005)
TITLE	The FANTOM Consortium, Riken Genome Exploration Research Group and
JOURNAL	Genome Science Group (Genome Network Project Core Group).
REFERENCE	The Transcriptional Landscape of the Mammalian Genome
AUTHORS	Science 309, 1559-1563 (2005)
TITLE	8 (bases 1 to 2371)
JOURNAL	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
REFERENCE	Fukuda, S., Furuno, M., Hasegaki, T., Hara, A., Hashizume, W.,
AUTHORS	Hayashida, K., Hayase, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
	Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
	Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
	Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
	Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
	Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
	Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
	Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
	Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
REFERENCE	Physical and Chemical Research (RIKEN), Laboratory for Genome
AUTHORS	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
	RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
	Kanagawa, 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
	URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
	Fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome
	Encyclopedia Project of Genome Exploration Research Group in Riken
	Genomic Sciences Center and Genome Science Laboratory in RIKEN.
	Division of Experimental Animal Research in Riken contributed to
	prepare mouse tissues.
	Please visit our web site for further details.
	URL: http://genome.gsc.riken.jp/
	Location/Qualifiers
FEATURES	1. .2371
source	/organism="Mus musculus"
	/mol_type="mRNA"
	/strain="C57BL/6J"
	/db_xref="FANTOM_DB:C230097E14"
	/db_xref="taxon:10090"
	/clone="C230097E14"
	/tissue_type="cerebellum"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="0 day neonate"
	437. .2290
CDS	/note="unnamed protein product; hypothetical protein
	(evidence: rscDS, ProCrest, decoder, NCBI CDS
	Predictor, Longest-ORF)
	putative"
	/codon_start=1
	/protein_id="BAC33536.1"
	/db_xref="GI:26339746"
	/translation="MGDMVTDPVLCTENKNLSQYTSRTKMSPSLSYQQVLCSSVP
	LSKNVGVFGVCTGSEIQSISYLDQELTTFPGPSLYESKSKAEKRLNITVALNC
	MNELLVLQRKNLLAQESVETQNLKLGSDMDHLOSCYAKLKEQLTSREMLGQERDR
	QKQNLRLHOLLKAEQDEKVKQNLNIASRTQNHDKRKEREYNKLRHLQVLVN
	KQDNIAVDLVYVGRADKKGKSWRTDEAGNEDEMYKILLNDEYKQKLMEAE
	LKKYLQMKKEMISLSPPKKPRAEADGTGTVAISDIEDSGELSDVSGLSCDT
	VREQNTSIRKQWRLKSHVKLDNQASKVHSEGNEDVTSRQDHEQETKLEIE
	RCKEMIRAKQQLLOQLATTCDDDTSLRLDCYLLERLKEEWTLFKEQKNFRER
	RSFTAAIRLGLERKAFEEERASWVKQFLNMTNFDHNSNPKLFSFSGSKPGSDP
	DNLLVHSRPROKLLHSVANGVPACTSKLTSLPASPTSDFRQTHSCVSEISVLN
	ITPESKPSPEVARESTDQKVSQSRPSREGCYSCGSAFRSAGHRRDLP"

## ORIGIN

**Alignment Scores:**

Pred. No.:	8.68e-303	Length:	2371
Score:	3138.00	Matches:	614
Percent Similarity:	99.4%	Conservative:	0
Best Local Similarity:	99.4%	Mismatches:	0
Query Match:	99.1%	Indels:	4
DB:	6	Gaps:	2

US-10-644-084-2 (1-615) x AK049080 (1-2371)

QY	1	MetGlyAspTyrMetThrValThrAspProValLeuLeuCysThrGluAsnLysAsnLeuSer	20
DB	437	ATGGGAGATTGGATGACTGTGACAGATCCAGTTCTGTGTACAGAAAAACAAAAATCTCTCT	496
QY	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
DB	497	CAATATACCTCAGAAACAAAGATGCTCCGTCCAGTTGTACTCCACAGCAAGTTCTGTGC	556
QY	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyValGlu	60
DB	557	TCTTCAGTACCTTTATCCAAAAACGTGATCGTGTCTTCGGTGTCTTCACAGAGAG	616
QY	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
DB	617	AACAVTTGAACAAAGTATTTCCTATCTTGATCAGGAGCTGACCACCTTCGGGTTTCCTTCC	676
QY	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu	100
DB	677	TTGTATGAAGAAATCCAAAGTAAGAGCAAGAGAGAAATTAATATATAGTCGCTGTTCTG	736
QY	101	AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal	120
DB	737	AACTGTATGAACAGCTGCTCGTCTCAGCGGAAGAACCTGTGTCGCCAGGAGACGGTG	796
QY	121	GluThrGlnAsnLeuLysGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys	140
DB	797	GAGACACAGAACTTGAAGCTGGCAGTGACATGACCACCTGCAGAGCTGTACTACGCCAA	856
QY	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg	160
DB	857	CTTAAGGAGCAGTTGGAAACGTCACGCGGAGATGATCGGGCTTCAAGAGAGACAGCG	916
QY	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal	180
DB	917	CAGCTGCAGTGCNAGAACAGAGGTTTGCATCAGCTCCTCGAAGAAATGAGAAAGTAGGTA	976
QY	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg	200
DB	977	CAAAAAATTACAAAAATATCATAGCCAGCGGGCTACTCAGTATAATCATGATGTGAAGAG	1037
QY	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys	220
DB	1037	AAGGAGCGTGAATATAATTAAGCTAAGAGCGCGCTGCATCAGCTCGTTATGAACAGAAG	1097
QY	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240
DB	1097	GATAAANAACATAGCCATCGATGTTTTAAATATATGTGGTGCAGCTGATGCAACACGAGC	1157
QY	241	SerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
DB	1157	TCATGGGAGCTGCAAAACAGAGCCAGGAATGAAGATCAGATGATGACAAAATTCGTGTG	1217
QY	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal	280
DB	1217	AATGATATTAGTACCGCCAGAACACGATCCTGTATGAGAACCGGAGCTGAAGAAGGTC	1277
QY	281	LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg	300
DB	1277	CTCCAGCAGATGAGAGAGGATGATCTCTCTCTGTCCTCAGAAGAGAGAGCCAGG	1337
QY	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly	320

Db	1337	GAAGAGCAGAGGACGGCA	CAGGCACTGTTGCTATCTCCGATATAGAAGATGACTCTGGG	1396
Qy	321	GlulSerArgAspSerValT	rpGlyLeuSerCysAspThrValArgGluGlnLeuThr	340
Db	1397	GAAC TGAGCAGACAGACG	GTGTGGGCGCTTTCTGTGCACACTGTGAGAGACAGCTGACA	1456
Qy	341	AsnSerIleArgIysGlnT	rpArgIleLeuIysSerHisValGlnIlyysLeuAspAsnGln	360
Db	1457	AACAGCATCAGAAACAG	TGGAGAAATTTTGAAGAATCATGTAGAAAACTCGATAACAA	1516
Qy	361	AlaSerLysValHisSer	GluGlyLeuAsnGluAspValIleSerArgGlnAspHis	380
Db	1517	GCITTCGAAGGTACACT	CAGAGGCGCTTAATGAGGAGGACGTCATCTCACGACAGACCAT	1576
Qy	381	GluGlnGlnThrGluLys	LeuGluLeuGluIleIleuArgCysIysGluMetIleLysAla	400
Db	1577	GAGCAAGAGACTCGAA	AACCTGGAGCTGGAGATTGAGCGGTGAAGAGATGATCAAGGCT	1636
Qy	401	GlnGlnGlnLeuLeuG	lnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu	420
Db	1637	CAGCAGCAGCTCTTA	----GAGCAGCTGGCCACCACCGTGTGATGATGACACCACTCTACTG	1693
Qy	421	LeuArgAspCysTyrLeu	LeuGluGluLysGluArgLeuLysGluGluTprThrLeuPhe	440
Db	1694	TTGCGAGACTGTTACT	TTGCTTGGAGAAAGAAAGCGCTTAAAGAGAGTGGACCTTTT	1753
Qy	441	LysGluGlnIlyssIys	AsnPheGluArgGluArgSerPheThrGluAlaIleArg	460
Db	1754	AAAGAGCAAAAAAGAA	ATTTTGAGAGAAAGCGAAGCTTTACAGAAAGTGCATTCGA	1813
Qy	461	LeuGlyLeuGluArgLys	AlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe	480
Db	1814	TTGGGGTTGGAGAGAA	AGGGCTTTGAAGACAGCGAGCCAGCTGGGTAAAGCAGCAGT	1873
Qy	481	LeuAsnMetThrAsnP	heAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe	500
Db	1874	TTAAACATGACGAACT	TTTGACCACCAAGACTCAGAAAACTGTGAAACTTTTTCAGTGCCTTC	1933
Qy	501	SerGlySer-----	SerAspProAspAsnLeuIleValHisSerArgProArgGln	517
Db	1934	TCAGGAAGTAAACCA	GGTTCGTATCCAGACAAATCTTTATAGTCCACTCACGGCCACGGCAA	1993
Qy	518	LysLysLeuHisSerVal	AlaAsnGlyValProAlaCysThrSerLysLeuThrLysSer	537
Db	1994	AAGAGCTACACAGTGT	GGCTAATGGGGTGCCAGCTTGACATCTAAAACTGACTAATCT	2053
Qy	538	LeuProAlaSerProSer	ThrSerAspPheArgGlnThrHisSerCysValSerGluHis	557
Db	2054	CTTCTCGCTCCACTCT	CTACTTCAGACTTTCCGCAGACACATTCATGTGTGCTGAACAC	2113
Qy	558	SerSerIleSerValLeu	AsnIleThrProGluGluSerIlyssProSerGluValAlaArg	577
Db	2114	AGTTTCATCATGTGCT	CGTAATATAACTCTCGAAGAAAGTAAACCAAGTAGGTGGAAGA	2173
Qy	578	GluSerThrAspGlnLys	TrpSerValGlnSerArgProSerSerArgGluGlyCysTyr	597
Db	2174	GAAGCAGCGATCAGA	AGTGGAGCGTGCAGTCGAGGCCAGCTCGCGGAGGGGTGCTAC	2233
Qy	598	SerGlyCysSerSerAla	PheArgSerAlaHisGlyAspArgAspLeuPro	615
Db	2234	AGCGGATGCTCTCGC	CTTTCAGAGCGGCTCACGGGACCGAGATGACTTACCT	2287

## RESULT 6

AY412491

**LOCUS**

## DEFINITION

SECRET

ACCESSION  
NUMBER

VERSION  
KEYWORDS

**KEYWORDS**  
**SOURCE**

**SOURCE**  
**ORGANI**

AY412491 1845 bp DNA linear GSS 12-DEC-2003  
Homo sapiens HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.

AY412491  
AY412491.1 GI:39768456  
GSS.  
Homo sapiens (human)  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Homo.

## REFERENCE

## AUTHORS

1 (bases 1 to 1845)  
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.

## TITLE

Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios

## JOURNAL

Science 302 (5652), 1960-1963 (2003)

## PUBMED

14671302

## REFERENCE

## AUTHORS

2 (bases 1 to 1845)  
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,  
Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,  
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,  
Adams, M.D. and Cargill, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA

This sequence was made by sequencing genomic exons and ordering  
them based on alignment.

## FEATURES

## source

1..1845  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
<1..>1845  
/locus\_tag="HCM4575"

## gene

## ORIGIN

## Alignment Scores:

Pred. No.: 18-267 Length: 1845  
Score: 2786.50 Matches: 538  
Percent Similarity: 93.2% Conservative: 35  
Best Local Similarity: 87.5% Mismatches: 41  
Query Match: 88.0% Indels: 1  
DB: 14 Gaps: 1

US-10-644-084-2 (1-615) x AY412491 (1-1845)

Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
Db 1 ATGGGAGATTGGATGACTGTTACAGATCCAGGTCTGCTTCAGAAAGCAAACTATCTCT 60  
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Db 61 CAATATACCTCGAAACCAAGATGCTCCATCAAGTTTATATACTACAGCAAGTGCTATGT 120  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 121 TCTTCATACCTTTATCGNAAAATATGTCACAGTTTTTTTCAGTGCCTTCTGACACAAAGAT 180  
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
Db 181 AATATTGAACACAGTATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTTCTCTCA 240  
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Db 241 TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGATTAATAATATAGTAGCTACTA 300  
Qy 101 AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
Db 301 AATTGTATGAATAGCTGCTGCTTTCAGCCGGAAGACCTTCTAGCTCAGGAAAATGTG 360  
Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db 361 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCTCTACAGAGCTGCTACTCAAAA 420  
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Db 421 CTTAAGGAACAACTGGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAAGAGACAGA 480

Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluVal 180  
Db CAGTTACAAATGTAGAACAGGAATTTGCATCAGCTACTAAGAAATGAGAAAGATGAGGTG 540  
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
Db CAAAAATACAAAATATCATTGCAAGTCGAGCTACTCAGTATAATCATCATATGAAGAGA 600  
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Db AAAGAGCGTGAATATAATAAACTGAAGAACGCTCTACATCAACTTGTATGAACCAAGAAA 660  
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db GATAAGAAAATAGCTATGACATTTTGAATTTATGTCGGAGAGAGCTGATGGAANAAGAGGC 720  
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Db TCCTGGAGGACTGGTAAAACTGAAGCCAGGAATGAAGATGAATGTATAAAATTTCTCTTG 780  
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
Db AATGATTATGAATATCGTCAGAAAACAAATCCTAATGGAAAATGCAAGAACTTAAGAAGTT 840  
Qy 281 LeuGlnGlnMetLysLysGluMetLysSerLeuLeuSerProGlnLysLysLysProArg 300  
Db CTTCAACAATGAAAAGGAAATGATTTCTCTCTTCTCCCAAAAAGAGAAACCTAGA 900  
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
Db GAAAGAGTAGATGATAGTACAGGAACCTGTT---ATTTCCGATGTTGAAGAAGATGCCGGG 957  
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
Db GAACTTAAGCAGAGAGATGATGGGACCTTCTCTGTGAACATGTGAGAGAGCGCTTACA 1017  
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
Db AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGTCATGTAGAAAAGCTTGATAACCAA 1077  
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380  
Db GTTTCAAAGGTACACCTGGAAGGTTTTAATGATGAAGATGTAATCTCAGCAAGACCAT 1137  
Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
Db GAAACAAGAACTGAAAACACTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAAACT 1197  
Qy 401 GlnGlnGlnLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
Db CAGCAACAGCTTTTACACGACGAGCTCGCTACTGCATATGATGATGATACCACTTCACTA 1257  
Qy 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440  
Db TTACGAGACTGTTATTTGTTGGAAGAAAGAACGCTCTCAAAGAAAGAAATGGTCCCTTTT 1317  
Qy 441 LysGluGlnLysLysAsnPheGluArgArgSerPheThrGluAlaAlaIleArg 460  
Db AAAGAGCAAGAAAAGAAATTTTGAAGGAGAGACGAGCTTTTACAGAGCCGCTATTCCG 1377  
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480  
Db CTGGGATTTGGAGAAAGGCACTTTGAAGAAAGAAAGAGCCAGTTGGTTAAAGCAGCAGTTT 1437  
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500  
Db CTAATAATGACTACCTTTTGACCAACAGAACTCAGAAAAATGTGAAAATTTTTCAGTGCCTTC 1497  
Qy 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
Db TCAGAGAGTCTGATTGGGACAATCTTATAGTGCACCTCAGGCGACCCGCAAGAGACCT 1557  
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540

```

|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1558 CACAGTGTCTTAATGGGCTCCAGTTTGCATGCTTAACCTACTAAATCTCTTCCTGCT 1617
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1618 TCACCTTCCACTTCAGACTTTTGCCAGACAGACGCTTCCTGCATATCTCAACATAGTTCAATC 1677
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1678 AATGTACTGAATATACTGCTGAAGAAATTTAAACCAATCAGGTTGGAGAGAATGTACA 1737
Qy 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGluCysTrpSerGlyCys 600
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1738 AATCAAAATGGAGTGTGGCATCAAGACTTGGATCACAGGAAGTTGCTATAGTGGATGC 1797
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1798 TCGTTGAGCTACACAAATCTCATGTAGAAAAGATGACTTACCT 1842

RESULT 7
LOCUS CR613292 2111 bp mRNA linear HTC 21-JUL-2004
DEFINITION Full-length cDNA clone CS0DL007Y108 of B cells (Ramos cell line)
Cot 25-normalized of Homo sapiens (human).
ACCESSION CR613292
VERSION 1
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 2111)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 2111)
Genoscope.
DIRECT SUBMISSION
SUBMITTED (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..2111
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL007Y108"
/tissue_type="B cells (Ramos cell line) Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.: 1.24e-267 Length: 2111
Score: 2786.50 Matches: 538
Percent Similarity: 93.2% Conservatve: 35
Best Local Similarity: 87.5% Mismatches: 41
Query Match: 88.0% Indels: 1
DB: 6 Gaps: 1

US-10-644-084-2 (1-615) x CR613292 (1-2111)

Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 206 ATGGAGATGGATGACTGTGTACAGATCCAGGTCTGTCTTCGAAAGCAAACTATCTCT 265

```

```

Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 266 CAATATACCTCAGAACAAAGATGTCTCCATCAAGTTTATATCTACACAGCAAGTGTATGT 325
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 326 TCTTCAATACCTTTATTCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAGAT 385
Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 386 AATATTGAACACAGTATCTCATATCTTGATCAGGAATTCAGTACTCTTTTGGTTTCTCTTCA 445
Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 446 TTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGAGTAAATATATAGTAGCTGACTA 505
Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 506 AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAAAATGTG 565
Qy 121 GluThrGlnAsnLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 566 GAGACACAGAAATTTGAGCTGGAGTGTATATGGACCATCTACAGAGCTGCTACTCAAAA 625
Qy 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 626 CTTAAGGAACAACTCGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAGAGACAGA 685
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 686 CAGTTTACAATGTAAAGAACAGGAATTTGCATCAGCTACTAAAGAATGAGAAAGATGAGGTG 745
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 746 CAAAAATTTACAAATATCATTCAGAGTCCAGCTACTCAGTATATCATGATATGAAGAGA 805
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 806 AAAGAGCGTGAATATAATAAACTGAAGAACCGCTCTACATCAACTTGTATTGAACAGAAA 865
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 866 GATAAGAAAAATAGCTATGACATTTTGAATTTATGTGCGGAGAGCTGATCGAAAAAGAGCG 925
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 926 TCCTGGAGGACTGGTAAACCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 985
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 986 AATGATTATGAATATCGTCAGAAACAAATCCTAATGGAATAATGCAGAACTTAAAGAGGTT 1045
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1046 CTTCAACAAAATGAAAGAGAAATGATTTCTCTTTCTCCCAAAAGAAAGAAAGAAAGAAAG 1105
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1106 GAAAGAGTAGATGATAGTACAGGAATCTGTT---ATTTCCGATGTTGAGAAAGATGCCGGG 1162
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1163 GAACCTAAGCAGACAGAGTATGTGGACCTTCTCTGTAAGCTGTGAGAGAGCAGCTTACA 1222
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1223 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAGTCTAGTAAAAAGCTGTGATAACCAA 1282
Qy 361 AlaSerLysValHisSerGluGluLeuAsnGluGluAspValIleSerArgGlnAspHis 380
|||||...|||||...|||||...|||||...|||||...|||||...|||||...
Db 1283 GTTTCAAAGGTACACCTCGAAGGTTTTTAATGATGAAGATGTAATCTCAGACAGACCAT 1342

```



```

Qy 381 GluGlnGluThrGluLysLeuGluGluLeuGluArgCysGluMetileLysAla 400
Db 1343 GAACAGAAACTGAAAACTCAGTTAGAAATTCAGCAGTGAAGAATGATTAACACT 1402
Qy 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThrThrSerLeu 420
Db 1403 CAGCAACAGCTTTTACAGCAGCAGCTCGTACTGTCATATGATGATACCACTTCACTA 1462
Qy 421 LeuArgAspCysTyLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
Db 1463 TTACGAGACTGTATTGTTGGAAGAAAAGGAACGTCTCAAGAAGAATGGTCCCTTTT 1522
Qy 441 LysGluGlnLysLysAspPheGluArgGluArgSerPheThrGluAlaAlaLeuArg 460
Db 1523 AAAGAGCAGAAAAGAAATTTTGAGAGGGAGAGACGAGCTTTACAGAGCCGCTATTGCG 1582
Qy 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnPhe 480
Db 1583 CTGGGATTGGAGAGAAGCAATTTGAAGAAGAAAGAGCAGTGTGTTAAAGCAGCAGTTT 1642
Qy 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1643 CTAAATATGACTACCTTTTGACCACAGAACTCAGAAAATGTGTAAGCAGCAGTTT 1702
Qy 501 SerGlySerSerAspProAspAsnLeuLeuValHisSerArgProArgGlnLysLysLeu 520
Db 1703 TCAGGAAGTTCGATTGGGACAACTTATATGTGCACTCGAGCAGCCGCAAGAAAGCGCT 1762
Qy 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1763 CACAGTGTCTAATGGGTCTCCAGTTGTCATGTCTAAACTTACTAAATCTCTTCTGCT 1822
Qy 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerLeu 560
Db 1823 TCACCTTCCACTTCAGACTTTTGCAGACACAGCTTCTGCATATCGAACATAGTTCAATC 1882
Qy 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgLysThr 580
Db 1883 AATGTACTGANTATACTGCTGAGGAAATTAACCAATCAGTTGGAGGAGATGTACA 1942
Qy 581 AspGlnLysTrpSerValGlnSerArgProSerArgGluGluGlyCysTrpSerGlyCys 600
Db 1943 AATCAAAATGGAGTGTGCATCAAGACCTGGATCACAGGAAGTTGCTATAGTGGATGC 2002
Qy 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Db 2003 TCCTTGAGCTACACAATTTCTATGTAGAAAAGATGACTTACCT 2047

RESULT 8
LOCUS BC035580
DEFINITION Homo sapiens synovial sarcoma, X breakpoint 2 interacting protein,
mRNA (CDNA clone IMAGE:5575710).
ACCESSION BC035580
VERSION BC035580.1 GI:22028261
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 3154)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Udén,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Rana,S.S., Loquellano,N.A., Peters,G.J.,
Armstrong,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

```

```

Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettaman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywicki,M.I., Skalska,U., Smalius,D.E.,
Schnorch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Mammalian Gene Collection Program Team
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3154)
NIH MGC Project
Direct Submission
Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@hgrl.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
McDowell,J., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Taurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 69 Row: n Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 51317360
This clone has the following problem: no 5' EST match.
FEATURES
Location/Qualifiers
1..3154
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5575710"
/tissue_type="Testis, embryonal carcinoma"
/clone_lib="NIH_MGC_92"
/lab_host="DH10B"
/note="vector: pCMV-SPORT6"
ORIGIN
Alignment Scores:
Pred. No.: 2,33e-267 Length: 3154
Score: 2786.50 Matches: 538
Percent Similarity: 93.2% Conservative: 35
Best Local Similarity: 87.5% Mismatches: 41
Query Match: 88.0% Indels: 1
DB: 6 Gaps: 1
US-10-644-084-2 (1-615) x BC035580 (1-3154)
Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysLeuSer 20
Db 86 ATGGAGATTGGATGACTGTTACAGATCCAGCTGCTGTTCCAGAAAGCAAACTATCTCT 145
Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 146 CAATATACCTCAGAAACAAGATGTTCCATCAAGTTTTATATCTACAGCAAGTGTATGT 205

```





available at <http://mips.gsf.de/projects/cdna/>.

FEATURES		Location/Qualifiers	
source	1..5658	/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="RZPD:DKFZp434C0515"	
		/db_xref="taxon:9606"	
		/clone="DKFZp434C0515"	
gene		/issue_type="testis"	
		/clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"	
		/dev_stage="adult"	
		/note="synovial sarcoma, X breakpoint 2 interacting protein"	
	1..5658	/gene="DKFZp434C0515"	
CDS	267..2111	/gene="DKFZp434C0515"	
		/codon_start=1	
		/product="hypothetical protein"	
		/protein_id="CA61373.1"	
		/db_xref="GI:6453459"	
		/db_xref="UniProtKB/TrEMBL:Q9UIX0"	
		/translation="MGDMVTVDPLGSSEKTIISQYTSSTKMSPSLSYQQVLCSSIP LSKNVHSPFACTEDNDTEQSIYLDQBLITFPFSPSLYEEKSGKETKRELNIIVALNC MNELVALORKNLADENVEQNLKLGSDMDHLQSCYSKLKSOLETSRREMILGQERDR QLOCKNRNLHQLLNKDEQVKQLNIIASRATQYNDHMKRERYNKLKERLHQLVMN KDKKIAMDILNYGRAGDKSWRTGKTEARNEDEMYKILLNDYFKQKILMENAE LKKVLQMKKEMISLSQKKPKPRVDSTGTVISDVEEDAGELSRRESMDLSCETV REQLTNSRKQWIRILKSHVEKLDNQVSKVHLEGFNEDVI SRQHEQTEKLELEIQ CKEMIKTOQLQOQLATAYDDTTSLRDLQVLLEEKRLKEWSLFPKEQKNPERER RSPTAAIIRLGLERKAFPERASWLKQOFLANWTFPDHONSENVKLFSFSGSSDWNIL IVHSQPOKKHSHVSNVSGPVCMSKLTSLPASPSDSFCQTRSCISEHSSINVLNITA BEIKPNQVGGERTNOKWSVASRPGSGEGCYSGCSLSYTNSHVEKDDLP"	
		ORIGIN	
	Alignment Scores:		
	Pred. No.:	5,86e-267	Length: 5658
	Score:	2786.50	Matches: 538
	Percent Similarity:	93.2%	Conservative: 35
	Best Local Similarity:	87.5%	Mismatches: 41
	Query Match:	88.0%	Indels: 1
	DB:	6	Gaps: 1
US-10-644-084-2 (1-615) x HSM801317 (1-5658)			
Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20	
Db	267	ATGGGAGATTGGATGACTGTTACAGATCCAGGTCTGCTCTCAGAAAAGCAAACTATCTCT 326	
Qy	21	GlnTyrThrSerGluThrIleMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40	
Db	327	CAATATACCTCAGAAACCAAGATGCTCCATCAAGTTTATATCTCACGCAAGTGTCTATGT 386	
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60	
Db	387	TCTTCAATACCTTTATCGAAAATGTGCACAGATTTTTCAGTGCCTTCTGCACAGAGAT 446	
Qy	61	AsnIleGluGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80	
Db	447	AATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTTTTGGTTTTCCTTCA 506	
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100	
Db	507	TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGATTAAATATATAGTAGCTACTA 566	
Qy	101	AsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120	
Db	567	AATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAGAGACCTTCTAGCTCAGGAAAATGTG 626	
Qy	121	GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140	
Db	627	GAGACACAGAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGTCTACTCAAAA 686	

Qy	141	LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160	
Db	687	CTTAAGGAACAACCTGGAAACCTCCAGGAGGAAATGATTGGCTTCAGGAAGACAGACAGA 746	
Qy	161	GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180	
Db	747	CAGTTACAACTGTAAGAACAGGAATTTGCATCAGCTACTAAAGATGAGAAACATCAGGTG 806	
Qy	181	GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200	
Db	807	CAAAAATACAAAATATCATTCGAAGTCGAGCTACTCAGTATAATCATGATATGAAGAGA 866	
Qy	201	LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220	
Db	867	AAAGAGCGTGAATATAATAAACTGAAGGAACGCTCATCATCACTTGTATTGAAACAGAAA 926	
Qy	221	AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240	
Db	927	GATRAAGAAAATAGCTATATGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAGGC 986	
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260	
Db	987	TCCTGGAGGACTGGTAAACTGAAGCCAGGAATGAAGATGAATAATATAAAATTTCTCTTG 1046	
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280	
Db	1047	AATGATTATGAATTCCTGCAGAAACAAATCCTTAATGGAATAATGCAGAACTTAAGAAGTT 1106	
Qy	281	LeuGlnGlnMetLysLysGluMetLysSerLeuLeuSerProGlnLysLysLysProArg 300	
Db	1107	CTTCAACAAATGAAAAGGAAATGATTCTCTCTTTCTCCCAAAAGAAAGAAACCTAGA 1166	
Qy	301	GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320	
Db	1167	GAAGAGTAGATGATAGTACAGGAAGTCTTATTTCCGATGTTGAAGAAAGATCCCGG 1223	
Qy	321	GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340	
Db	1224	GAATTAAGCAGAGAGATGTATGGGACCTTCTCTGTGAAACTGTGAGAGAGAGAGCTTACA 1283	
Qy	341	AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360	
Db	1284	AACAGCATCAGAAACACAGTCGAGAATTTTGAAGATCATGTAGAAAAGCTTGATAACCAA 1343	
Qy	361	AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380	
Db	1344	GTTTCAAGAGGTACACCTCGAAGGTTTAAATGATGAAGATGTAATCTCAGCAAGACCAT 1403	
Qy	381	GluGlnGluThrGluLysLeuGluLeuGluArgCysLysGluMetIleLysAla 400	
Db	1404	GAACAGAAACTGAAAAACTCGAGTAGAATTCAGCAGTCTAAAGAAATGATTAAACT 1463	
Qy	401	GlnGlnGlnLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420	
Db	1464	CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCTATGATGATGATGATGATGATGAT 1523	
Qy	421	LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440	
Db	1524	TTACGAGACTGTTATTTGTTGGAAGAAAGAACCTCTCAAAGAAAGAAATGGTCCCTTTT 1583	
Qy	441	LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg 460	
Db	1584	AAAGACAGAAAAGAAATTTTGAAGGAGAGAGACGACGAGCTTTTACAGAGCCGCTATTCCG 1643	
Qy	461	LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnPhe 480	
Db	1644	CTGGATTGGAGAGAGAGGCATTTGAAGAAAGAGAGCCAGCTGGTTGTTAAAGCAGCAGTTT 1703	
Qy	481	LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500	
Db	1704	CTAAATATGACTTACCTTTTGACCCACCAAGAACTCAGAAAATGTGAAAATTTTTCAGTGCCTTC 1763	

QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520  
 Db 1764 TCAGGAAGTCTCTGATTGGGCAATCTTATAGTGCACCTCGAGGCGCGCAAAAAGACCT 1823  
 QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540  
 Db 1824 CACAGTGTGCTAATGGGTCTCCAGTTTGATGTCTAAACTTACTAAATCTCTCTGCT 1883  
 QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSile 560  
 Db 1884 TCACCTTCCACTTCAGACTTTTCCGACACAGCTTCTCGCATATCTGAACATAGTCAATC 1943  
 QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580  
 Db 1944 AATGTACTGAATATTAATCTCTGAAGAAATTAACCAAAATCAGGTGGAGAGAACGTACA 2003  
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600  
 Db 2004 AATCAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGGAAGTTGCTATAGTGGATGC 2063  
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615  
 Db 2064 TCCTTGAGCTACAAAATTCATCTAGTAAAAAGATCACTTACCT 2108

## RESULT 10

AY412492

## LOCUS

DEFINITION Pan troglodytes HCM4575 gene, VIRTUAL TRANSCRIPT, partial sequence,  
 genomic survey sequence.

ACCESSION AY412492

VERSION AY412492.1 GI:39768457

KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)

ORGANISM

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Pan.

1 (bases 1 to 1845)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,

Fertig, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Inferring nonneutral evolution from human-chimp-mouse orthologous

gene trios

Science 302 (5652), 1960-1963 (2003)

14671302

2 (bases 1 to 1845)

Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,

Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,

Fertig, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

Adams, M.D. and Cargill, M.

Direct Submission

Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,

Rockville, MD 20850, USA

This sequence was made by sequencing genomic exons and ordering

them based on alignment.

Location/Qualifiers

FEATURES

source

1..1845

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

&lt;1..&gt;1845

/locus\_tag="HCM4575"

## ORIGIN

Alignment Scores:

Pred. No.: 8,13e-264 Length: 1845

Score: 2747.50 Matches: 531

Percent Similarity: 92.0% Conservative: 35

Best Local Similarity: 86.3% Mismatches: 48

Query Match: 86.8% Indels: 1

DB: 14 Gaps: 1

## gene

source

1..1845

/organism="Pan troglodytes"

/mol\_type="genomic DNA"

/db\_xref="taxon:9598"

&lt;1..&gt;1845

US-10-644-084-2 (1-615) x AY412492 (1-1845)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
 Db 1 ATGGGAGATTGGATGACTGTTACAGATCCAGGCTCTGCTTCAGAAAGCAAACTATCTCT 60  
 QY 21 GlnTyThrSerGluThrLysMetSerProSerSerLeuTySerGlnGlnValLeuCys 40  
 Db 61 CAATATACCTCGAAACAAAGATGCTCCATCAAGTTNNNACTCACAGCAAGTGTATGT 120  
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
 Db 121 TCTTCAATACCTTTATCGAAATAATGTCACAGTTTTTTCAGTCCCTTCTGCACAGAGAT 180  
 QY 61 AsnIleGluGlnSerIleSerTyLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
 Db 181 AATATTGAACACAGATATCTCATATCTTGATCAGGAATTCAGTACTTTTGGTTTTCTTCA 240  
 QY 81 LeuTyGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
 Db 241 TTATATGAAGAAATCCAAAGGTAAAGAGACAAAGAGAGAGTTAAATATAGTAGTGCTACTA 300  
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
 Db 301 AATTGTATGAATGAGCTGCTGTGCTTCAGCGAAGAACCTTCTAGTCAGGAAAATGTG 360  
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLys 140  
 Db 361 GAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAA 420  
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 Db 421 CTTAAGGAAACAACCTGGAAACCTCCAGGAGGGAATGATTGGGCTTCAGGAAAGAGACAGA 480  
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180  
 Db 481 CAGTTACAATGTAAAGAACACAGGAATTTGCATCAGCTACTAAAGAAATCAGAAAGATAGG 540  
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyAsnHisAspValLysArg 200  
 Db 541 CAAAATTTACAAAATATCATTCGAACTCGAGCTACTCAGTATAATCATGATATGNNNNN 600  
 QY 201 LysGluArgGluTyAsnLysLeuLysGluArgLysGlnHisGlnLeuValMetAsnLysLys 220  
 Db 601 AAAGAGCGTGAATATAATAANNNGNNNAACGCTCTACATCAACTCTTGTATGAACAAGAA 660  
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyValGlyArgAlaAspGlyLysArgGly 240  
 Db 661 GATAAGAAATAGCTATGGACATTTTGAATTAATGTCGGGAGAGCTGATGGAAGAGAGGC 720  
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyLysIleLeuLeu 260  
 Db 721 TCCTGGAGGACTGGTAAACCTCAAGCCAGGAATGAAGATGAATGTATATAAATTCCTTCT 780  
 QY 261 AsnAspTyThrGluTyArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
 Db 781 AATGATTATGAATATCGTCAAGAAACAAATCCTTAATGGAAAAATGCAGAACTTAAGAAG 840  
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
 Db 841 CTTCAACAAATGAAAAGGAAATGATTTCTCTCTTCTCCCAAAAAGAGAAACCTCTAGA 900  
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320  
 Db 901 GAAAGAGTAGATGATAGTACAGGAATCTGTT---ATTCTGATGTTGAGAAAGATGCCGGG 957  
 QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
 Db 958 GAACTAAGCAGACAGAGATGATGGGACCTTTCTCTGTAACCTGTGAGAGAGCAGCTTACA 1017  
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
 Db 1018 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAANAAGTCATGTAGAAAAAATTGATAACCAA 1077



```

QY 244 ThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyr 263
Db 422 ACTGACAAACAGAGCCAGGAGATGAGATGATGATGATGATGATGATGATGATGATGAT 481
QY 264 GluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGln 283
Db 482 GAGTACCGCCAGAGACAGATCCTGATGGAGAACCGGAGCTGAAGAAGGTCTCCAGCAG 541
QY 284 MetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAla 303
Db 542 ATGAAGAAGGAGATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 601
QY 304 GluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSer 323
Db 602 GAGACGGCACAGCAGCTGTTGCTATCTTCGATATAGAGATGACTCTGGGAACTGAGC 661
QY 324 ArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIle 343
Db 662 AGACAGAGCGTGTGGGCGCTTTCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATC 721
QY 344 ArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLys 363
Db 722 AGAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAAATCTCGAATACCAAGCTTCCAAG 781
QY 364 ValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHisGluGlnGlu 383
Db 782 GTACACTCAGAGGCGCTTAATGAGGAGGAGCGTCATCTCAGCAACAGACCATGAGCAAGAG 841
QY 384 ThrGluLysLeuGluLeuGluIleGluArgCysLys 395
Db 842 ACTGAGAAACTGGAGCTGGAGATTGAGCGGTGAAA 877

RESULT 12
DV893190
LOCUS DV893190
DEFINITION DV893190.1 GI:82827419
ACCESSION DV893190
VERSION DV893190.1
KEYWORDS EST.
SOURCE Bos taurus (cattle)
ORGANISM Bos taurus

REFERENCE
AUTHORS Moore, S., Alexander, L., Brownstein, M., Guan, L., Lobo, S., Meng, Y.,
Tasaguchi, M., Wang, Z., Yu, J., Prange, C., Schreiber, K., Shenmen, C.,
Wagner, L., Baia, M., Barbazuk, S., Barber, S., Babakiff, R.,
Beland, J., Chun, E., Del Rio, L., Gibson, S., Hanson, R.,
Kirkpatrick, R., Liu, J., Matsuo, C., Mayo, M., Santos, R. R., Stott, J.,
Tsai, M., Wong, D., Siddiqui, A., Holt, R., Jones, S. J. and Maria, M. A.
Bovine Genome Sequencing Program: Full-length cDNA Sequencing
Unpublished (2005)
TITLE Contact: Robert Kirkpatrick
JOURNAL Canada's Michael Smith Genome Sciences Centre
COMMENT BC Cancer Agency
Suite 100, 570 West 7th Avenue, Vancouver, British Columbia,
Canada, V5Z 4S6
Tel: 1-604-707-5900 x5406
Fax: 1-604-876-3561
Email: robertk@bcgsc.ca
Insert length: 903 Std Error: 0.00
Plate: LB02726 row: C column: 24
High quality sequence stop: 903.
Location/Qualifiers
1..903
/organism="Bos taurus"
/mol_type="mRNA"
/strain="L1 Hereford"
/db_xref="taxon:9913"
/clone="IMAGE:8315858"
/sex="female"

```

```

/tissue_type="Basal ganglia"
/dev_stage="8.5 months old calf"
/lab_host="E. coli DH10B T1 Phage resistant"
/clone_lib="GC_BGC-27"
/note="Vector: pExpress 1; Site 1: Blunt (5' end of cDNA);
Site 2: NotI (3' end of cDNA)"

ORIGIN
Alignment Scores:
Pred. No.: 6,91e-131 Length: 903
Score: 1416.50 Matches: 277
Percent Similarity: 96.3% Conservative: 13
Best Local Similarity: 92.0% Mismatches: 10
Query Match: 44.8% Indels: 1
DB: 10 Gaps: 1

US-10-644-084-2 (1-615) x DV893190 (1-903)
QY 78 PheProSerLeuTyrGluSerLysSerLysGluAlaLysArgGluLeuAsnIleVal 97
Db 3 TTTCCTTCATTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGGATTTAAATATAGTT 62
QY 98 AlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGln 117
Db 63 GCTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 122
QY 118 GluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCys 137
Db 123 GAAATGTGGAACACAGAAATTTGAAACTGGGAGCGATATGACCATCTACAGAACTGC 182
QY 138 TyrAlaLysLeuLysGluGlnLeuThrSerArgArgGluMetIleGlyLeuGlnGlu 157
Db 183 TATGCAGAACTTAAAGGACAGCTGGAAACCTCCAGGAGAGAAATGATGCTCTTCAGGAA 242
QY 158 ArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLys 177
Db 243 AGAGACAGACAAATTAATGCAAGATAGGAATTTGCATCAGCTACTGAAAATGAGAAA 302
QY 178 AspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAsp 197
Db 303 GATGAGGTGCAGAAATTAACAAATATCATTGCAAGTCGAGCTACTCAATATAATCATGAT 362
QY 198 ValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMet 217
Db 363 ATGAAGAGAGAAAGAGCGTGAATATATAATAAATAAAGGAAACGTTTACATCAACTTGTATG 422
QY 218 AsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGlyVargAlaAspGly 237
Db 423 AACAAAGAGGATTAATAAATTTCTATGGAAGTTTTAAATATTACGTGGGAGAGCTGATGG 482
QY 238 LysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLys 257
Db 483 AAAAGAGGCTCTCGAGGAGCGGTAAACCGAAGCGAGCAATGAAGATGAATGTATATAA 542
QY 258 IleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeu 277
Db 543 ATTCTCTTGAATGATTATGAATATCGTCAGAAACAAATCCTAATGAAAAATCTGCAACTT 602
QY 278 LysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLys 297
Db 603 AAGAAGGTTCTTCAGCAATGAAAAGGAAATGATTTCTCTCTCTCTCTCTCTCTCTCTCT 662
QY 298 LysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAsp 317
Db 663 AAACCTAGAGAAAGAGCAGATGATAGTACAGGAAGTGT--ATCTCTGATATTGAGAGAG 719
QY 318 AspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGlu 337
Db 720 GATGCTGGGAACTGAGTAGAGAGAGTATATGGGACCTTTCTCTGTCGAGACTGTGAGAGAG 779
QY 338 GlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeu 357
Db 780 CAGCTTACCAATAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAAGTCACGTAGAAAAACTT 839

```

QY 358 AspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArg 377  
 |||||  
 Db 840 GATAACCAAGTTTCAAAAGTACACTTAGAAGGTTTAAACGATGAGATGATATCTCACGA 899

QY 378 Gln 378  
 ||||  
 Db 900 CAA 902

RESULT 13  
 CN539511 804 bp mRNA linear EST 29-APR-2004  
 UI-W-HUO-cgu-n-01-0-UI.r1 NIH\_BMAP\_HUO Mus musculus cDNA clone  
 IMAGE:30667224 5', mRNA sequence.

ACCESSION CN539511

VERSION CN539511.1 GI:46867667

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 804)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. .804  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30667224"  
 /tissue\_type="whole eye"  
 /dev\_stage="newborn( 1, 5, 15 days )"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_HUO"  
 /note="Organ: Eye; Vector: pYX-Asc; Site\_1: EcoR I;  
 Site\_2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATNATTAGC. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

ORIGIN

Alignment Scores:  
 Pred. No.: 1,05e-125 Length: 804  
 Score: 1364.00 Matches: 265  
 Percent Similarity: 99.3% Conservative: 0  
 Best Local Similarity: 99.3% Mismatches: 2  
 Query Match: 43.1% Indels: 0  
 DB: 8 Gaps: 0

US-10-644-084-2 (1-615) x CN539511 (1-804)

QY 190 ArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLys 209  
 |||||  
 Db 2 CGGCTACTCTAGTATAATCATGATGTGAAGAGGAGCGCTGAATAATAATAAGCTAAAG 61

QY 210 GluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeu 229  
 |||||  
 Db 62 GAGCGCTGTCATCAGCTCGTTATGAACAGAAGGATAAAAACATAGCCATGGATGTTTTA 121

QY 230 AsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAla 249  
 |||||  
 Db 122 AATTATGTGGGTGCGAGCTGATGGCAACGAGGCTCATGGAGACTTGACAAAACAGAAGCC 181

QY 250 ArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGln 269  
 |||||  
 Db 182 AGGAATGAAGATGAGATGTACAAAATTCTGTTGAATGATTATAGTACCGCCAGAAGCAG 241

QY 270 IleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMetIle 289  
 |||||  
 Db 242 ATCTGTATGGAGAACGCGGAGCTGAAGAAGATCTCTCCAGCAGATGAAGAAGAGATGATC 301

QY 290 SerLeuLeuSerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309  
 |||||  
 Db 302 TCTCTCTGTCCTCTCAGAGAAGAAGCCCGAGGGAAGAGCAGGACGCGACAGGCACT 361

QY 310 ValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGly 329  
 |||||  
 Db 362 GTTGCTATCTCCGATATAGAAGATGACTCTCGGGAACCTGACAGAGACAGCGTGTGGGCG 421

QY 330 LeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIle 349  
 |||||  
 Db 422 CTTTCTGTGACACTGTGAGAGAGCAGCTGACAAACAGCATCAGGAAACAGTGGGAAT 481

QY 350 LeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeu 369  
 |||||  
 Db 482 TTGAAAAGTCTATGTAGAAAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCT 541

QY 370 AsnGluGluAspValIleSerArgGlnAspHisGluGlnGlnThrGluLysLeuGluLeu 389  
 |||||  
 Db 542 AATGAGGAGGAGCGTCATCTCACGACAAGACCATGAGCAAGAGACTGAGAAACTGGAGCTG 601

QY 390 GluIleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLysGlnGlnLeu 409  
 |||||  
 Db 602 GAGATTGAGCGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACACAGCAGCTG 661

QY 410 AlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGlu 429  
 |||||  
 Db 662 GCCACCGCTGTGATGATGACACCACCTCCTCTGTGCGAGACTGTTACTTCTTGGAGAA 721

QY 430 LysGluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArg 449  
 |||||  
 Db 722 AAGGAACGCTTAAAGAAGAGTGGACCCCTTTNTAAAGAGCAAAAAAGAAATTTTGAGAGA 781

QY 450 GluArgArgSerPheThrGlu 456

Db 782 GAAAGGCGAAGCTTTACAGAA 802

RESULT 14

CX562093

LOCUS UI-M-FC0-ctz-g-10-0-UI.r1 NIH\_BMAP\_F00 Mus musculus cDNA clone

DEFINITION IMAGE:30699705 5', mRNA sequence.

ACCESSION CX562093

VERSION CX562093.1 GI:57589122

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

TITLE Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 834)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

Seq primer: pYX-5.

## FEATURES

Location/Qualifiers

source

1..834  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30699705"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP FCO"  
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonafido, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is TGAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

Alignment Scores:  
Pred. No.: 7,076-125 Length: 834  
Score: 1356.00 Matches: 271  
Percent Similarity: 98.2% Conservative: 1  
Best Local Similarity: 97.8% Mismatches: 5  
Query Match: 42.8% Indels: 1  
DB: 9 Gaps: 0

US-10-644-084-2 (1-615) x CX562093 (1-834)

QY 91 LysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGln 110  
||| :|||  
Db 4 AAGAGAGAATAAATAGTCGCTGTTCTGAACGTGTATGAACGAGCTGCTGCTTCAG 63  
QY 111 ArgLysAsnLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerArg 130  
||| :|||  
Db 64 CGGAAGAAGCTGCTGCCAGGAGACGCGTGGAGACACAGAACTTGAAGCTGGGCAATGAC 123  
QY 131 MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArg 150  
||| :|||  
Db 124 ATGGACCACCTGCAGAGCTGTACGCCAAACTTAAAGAGCAGTTGGAAACGTCACGCCG 183  
QY 151 GluMetIleGluGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHis 170  
||| :|||  
Db 184 GAGATGATCGGGCTTCAAGAGAGAGACAGCGAGCTGCAGTGCAGCAAGACAGGAGTTTGCAT 243  
QY 171 GlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg 190  
||| :|||  
Db 244 CAGCTCCTGAAGAATGAGAAAGATGAGTACAAAATTTACAAAATATCATATGACCGCCG 303  
QY 191 AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGlu 210  
||| :|||  
Db 304 GCTACTCAGTATATCATGATGTGAAGAGAGAGGAGCGGTGAATATATTAAGCTAAGGAG 363

QY 211 ArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsn 230  
||| :|||  
Db 364 CGCTGTCATCAGCTCGTTATGAACAAGAGGATAAAAAATAGCCATGGATGTTTAAAT 423  
QY 231 TyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArg 250  
||| :|||  
Db 424 TATGTGGTCTGAGCTGATGGCAACGAGGCTCATGGAGGACTGACAAAACAGAAAGCCAGG 483  
QY 251 AsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270  
||| :|||  
Db 484 AATGAAGATGAGATGTACAAATTTCTGTTGAATGATTATGATACCCCGCAGAAGCAGATC 543  
QY 271 LeuMetGluAsnAlaGluLeuLysValLeuGlnGlnMetLysLysGluMetIleSer 290  
||| :|||  
Db 544 CTGATGGAGAACGCGAGCTGAAGAAGGTCCTCCAGCAGATGAAGAAGGAGATGATCTCT 603  
QY 291 LeuLeuSerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310  
||| :|||  
Db 604 CTCCTGTCTCTCAGAAAGAA-CCAGGGAAGAGCAGAGGACGCGCAGGACCTGTT 662  
QY 311 AlaIleSerAspIleGluAspSerGlyGluLeuSerArgAspSerValTrpGlyLeu 330  
||| :|||  
Db 663 GCTATCTCCGATATAGAACATGACTCTGGGAACTGAGCAGACAGACGCTGTGGGCTT 722  
QY 331 SerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeu 350  
||| :|||  
Db 723 TCTGTGACACTGTGAGAGCAGCTGACANACAGCATCAGGAAACAGTGGAGATTTTG 782  
QY 351 LysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGlu 367  
||| :|||  
Db 783 ANAAGTCATGTAGAAAAAATCGATAACCAAGCTTCGAGGTACACTCAGAG 833

CF949831 840 bp mRNA linear EST 20-NOV-2003  
UI-M-H30-cmt-d-22-0-UI.r1 NIH\_BMAP\_H30 Mus musculus cDNA clone  
IMAGE:30632061 5', mRNA sequence.  
CF949831  
EST.  
CF949831.1 GI:38465700  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 840)  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.

Location/Qualifiers  
1..840  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30632061"  
/tissue\_type="Upper Head"  
/dev\_stage="9.5 and 10.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_H30"  
/note="Organ: Head; Vector: pYX-Asc; Site\_1: EcoR I;

Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAATCTGAT. This library was created for the University of Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH)."

## ORIGIN

## Alignment Scores:

Pred. No.:	3,216-124	Length:	840
Score:	1349.50	Matches:	263
Percent Similarity:	98.9%	Conservative:	0
Best Local Similarity:	98.9%	Mismatches:	2
Query Match:	42.6%	Indels:	1
DB:	5	Gaps:	1

US-10-644-084-2 (1-615) x CF949831 (1-840)

Qy	350	LeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeu	369
Db	1	TTGAAAGTCATGTAGAAAACCTCATTAACCAAGCTTCGAGGTACACTCAGAGGGCCTT	60
Qy	370	AsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLeu	389
Db	61	AAATGAGGAGGACGTCTATCTCAGCAACAAGACCATGAGCAAGACTGAGAACTGGAGCTG	120
Qy	390	GluIleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuGlnGlnGlnLeu	409
Db	121	GAGATTGACGGTGTAAAGAGATGATCAAGGCTCAGCAGCAGCTCTTA---CAGCAGCTG	177
Qy	410	AlaThrThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGlu	429
Db	178	GCCACCACGTGTGATGATGACACCACTCCTGTTGCGAGACTGTTACTTGTCTGGAGAA	237
Qy	430	LysGluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArg	449
Db	238	AAAGGAACGCTTAAAGAGAGTGGACCCCTTTTAAAGAGCAAAAAGAAATTTTGAGAGA	297
Qy	450	GluArgArgSerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGlu	469
Db	298	GAAGGCGGAAGCTTTACAGAGCTGCCATTCGATTGGGGTTGGAGAGAAAGCGCTTGAA	357
Qy	470	GluGluArgAlaSerTrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGln	489
Db	358	GAAGAGCGAGCCAGCTGGGTAAAGCAGCAGTTTTAAACATGACGAACTTTGACCACCAG	417
Qy	490	AsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeu	509
Db	418	AACTCAGAAAATGTGAAACTTTTCAGTGCCTTCTCAGGAAGTCTCTGATCCACACAATCTT	477
Qy	510	IleValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAla	529
Db	478	ATAGTCCACTCACGGCCACGGCAAAAGAGCTACACAGTGTGGCTTAATGGGGTGCAGCT	537
Qy	530	CysThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGln	549
Db	538	TGCACATCAAAACTGACTAAATCTCTTCTGCTCACCCTTCTACTTCAGACTTTCGCCAG	597
Qy	550	ThrHisSerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGlu	569
Db	598	ACACATTTCATGTGTCTGAAACACACAGTTCATCAGTGTGCTGAAATATAACTCTCTGAAGA	657
Qy	570	SerLysProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArg	589
Db	658	AGTAAACCAAGTGAGGTTTGCAGAGAAAGCAGCGGATCAGAAAGTGGAGCGTCGAGG	717

Qy	590	ProSerSerArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGly	609
Db	718	CCCAGCTCGCGGAGGGGTGCTACAGCGGATGCTCTCGGNCCTTCANGAGCGCTCACCGN	777
Qy	610	AspArgAspAspLeuPro	615
Db	778	GACCGAGATGACTTACCT	795

Search completed: June 14, 2006, 01:45:29  
Job time : 7935 secs





Db 433 TACTTTTGGTTTCTCTTCATTATATGAAGATCCAAAGGTAAAGAGACAAAGAGAGAGTT 492  
Qy 361 AAATATAGTCTGCTGTTCTGAACCTGTATGAACGAGCTGCTGCTGCTTTCAGCGGAAGAACCT 420  
Db 493 AAATATAGTAGCTGTACTAAATTATGAATAGAGTCTGCTGCTTTCAGCGGAAGAACCT 552  
Qy 421 GCTGCCAGGAGAGCGTGGAGACACAGAACTTCAAGCTGGGCGAGTGACATGGACCACT 480  
Db 553 TCTAGCTCAGGAAATGTGGAGACAGAAATTTGAAGCCGGGAAGTATATGGACCATCT 612  
Qy 481 GCAGAGCTGTACGCCAACTTAAGGAGCAGTTTGAACCGTCCAGCGGGAGATGATCGG 540  
Db 613 ACAGAGCTGTACTCAAACTTAAGAACACTCGAAACCTCCAGGAGGGAATGATGG 672  
Qy 541 GCTTCAAGAGAGACAGGCGAGCTGCGAGTGOAAGAACAGGAGTTTGCATCAGCTCTGAA 600  
Db 673 GCTTCAGAAAGACAGACAGAGTTTACAAATGTAAGAACAGGAATTTGCATCAGCTACTAAA 732  
Qy 601 GAATGAGAAAGATCAGGTACAAAATTTACAAAATATCATAGCCAGCGGGCTACTCAGTA 660  
Db 733 GAATGAGAAAGATCAGGTGCAAAAATTTACAAAATATCATTTGCAAGTCAAGCTACTCAGTA 792  
Qy 661 TAATCATGATGTGAAGAGGAAGCGGTGAATATAAAGCTAAAGGAGCGCCCTGCATCA 720  
Db 793 TAATCATGATGTGAAGAGGAAGCGGTGAATATAAAGCTAAAGGAGCGCTTACATCA 852  
Qy 721 GCTCGTTATGAACGAAGAGTAATAAATCATAGCCATGATGTTTTAAATTTATGTGGGTG 780  
Db 853 ACTTGTATGAACGAAGAGTAATAAATAGCTATGACATTTTGAATTTATGTGCGGAG 912  
Qy 781 AGCTGATGGCAACCGAGCTCATCGGAGCTGACAAACAGAACCGAGGAATGAAGATGA 840  
Db 913 AGCTGATGGAAAGAGAGCTCTCGAGGACTGTGTAATAAAGCTGAAGCGAGGAATGAAGATGA 972  
Qy 841 GATGTACAAAATCTGTGTAATGATTTAGTATCCCGCAGAGCAGATCTCTGATGGAGAA 900  
Db 973 AATGTATAAATCTCTTGAATGATTAATATCGTCAGAAACAAATCTCTTAATGGGAAA 1032  
Qy 901 CGCGAGCTGAAGAGGTCTCTCAGCAGATGAAGAGAGATGATCTCTCTCTGCTCC 960  
Db 1033 TGCAGAACTTAAGAGGTCTCTCAACAAATGAAAGAGGAATGATTTCTCTCTCTCTCC 1092  
Qy 961 TCAGAAAGAAAGCCAGGGAAAGAGCAGAGCGGACAGCGCACTGTGTCTATCTCGA 1020  
Db 1093 CCAAGAGAGAAACCTAGAGAAAGATGATGATAGTACAGAACTGT---TATTTCCGA 1149  
Qy 1021 TATGAAGATGACTCTGCGGAACTGAGCAGAGACAGCGGTGTGGGCGCTTCTCTGTGACAC 1080  
Db 1150 TGTGAAGAGATGCGCGGGAACTAAGCAGAGAGATATGTGGGACCTTTCTCTGTGAAC 1209  
Qy 1081 TGTGAGAGAGCAGCTGACAAAACAGCATCAGGAAACAGTGGAGAAATTTTGAAGTCAATG 1140  
Db 1210 TGTGAGAGAGCAGCTTACAAAACAGCATCAGAAACAGTGGAGAAATTTTGAAGTCAATG 1269  
Qy 1141 AGAAAACTCGATAACCAAGCTTCAAGGTACACTCAGAGGCGCTTAATAGGAGGACGT 1200  
Db 1270 AGAAAGCTGATAACCAAGTTTCAAGGTACACTCGAAGGTTTAAATGATGAAGATG 1329  
Qy 1201 CATCTCAGCAAGACCATAGCAGAGACTGAGAAACTCGAGCTGGAGATTTGAGCGGTG 1260  
Db 1330 AATCTCAGCAAGACCATGAACAAAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTG 1389  
Qy 1261 TAAAGAGATGATCAAGGCTCAGCAGCAGCTTTACAGCAGCAGCTGGGCCACCACTGTGA 1320  
Db 1390 TAAAGAAATGATTAAGAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGATATGA 1449  
Qy 1321 TGATGACACCACTCAGCTGTTGCGAGACTGTACTTGTCTGGAAGAAAGAAAGCAAGCTTAA 1380  
Db 1450 TGATGATACCACTTCACTATTACGNGACTGTATTGTTGGAAGAAAGAAAGCTCTCAA 1509  
Qy 1381 AGAAGAGTGGACCTTTTAAAGAGCAAAAAAGAAATTTTGAAGAGAAAGCGCAAGCTT 1440  
Db 1510 AGAAGATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAAGAGGAGAGACGAAGCTT 1569

Qy 1441 TACAGAGCTGCCATTGATTGGGTTGGAG----- 1471  
Db 1570 TACAGAGCCGCTATTTCGCTGGGATTTGGAGATTTGGCTTCTCTGCAAGTGTTTTACA 1629  
Qy 1472 ----- 1471  
Db 1630 CAAGAAATCTCTTTTGGGTATCCATTCATCCAGGTCAGTAGGTGTGGTGTGAGAAG 1689  
Qy 1472 -----AGAAAGGCGTTTGAAGAGAGGAG 1496  
Db 1690 ACAGTGCACCAAGACTATATTTCGCTGTTTAAAAAAGAAAGGCAATTTTGAAGAAAGAGAG 1749  
Qy 1497 CCAGCTGGGTAAAGCAGCAGTTTTTAAACATGACGAACCTTTGACCACACAGAACTCAGAAA 1556  
Db 1750 CCAGTTGGTTAAAGCAGCAGTTTCTTAATATGACTACCTTTGACCACACAGAACTCAGAAA 1809  
Qy 1557 ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAAGTTTCTGATCCAGACAATCTTTATAGTCCACT 1616  
Db 1810 ATGTGAAACTTTTTCAGTGCCTTCTCAGGAAAGTTTCTGATTGGGACAACTTTATAGTGCAC 1869  
Qy 1617 CAGGCCACGGCAAAAGAGCTACAGAGTGTGGCTAATGGGGTCCAGCTTGCACATCAA 1676  
Db 1870 CGAGGAGCGCGCAAAAGAGCCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCTA 1929  
Qy 1677 AACTGACTAAATCTCTTCTGCTCCTCACCTTCTCAGACTTTTCGCGCAGACACATTCAT 1736  
Db 1930 AACTTACTAAATCTCTTCTGCTCCTCACCTTCCACTTCAGACTTTTGCAGACAGCTTCTCT 1989  
Qy 1737 GTGTGTCTGAACAGTTCCATCAGTGTGTGTAATATACTCTCTGAAGAAGTAAACCAA 1796  
Db 1990 GCATATCTGAACATAGTTCAATCAATGCTACTGAATATACTGCTGAAGAAATTAACCAA 2049  
Qy 1797 GTAGGTTGCAAGAGAAAGCAGATCAGAAAGTGGAGCTGAGCTGAGGCGCCAGCTCGC 1856  
Db 2050 ATCAGGTTGGAGGAGAAAGTCAAAATCAAAAATGGAGTGTGGCGTCAAGACCTGGATCAC 2109  
Qy 1857 GGAGGGGTCTACAGCGGATGCTCTCGGCTTCAGGAGCGCTCACGGGACCGAGATG 1916  
Db 2110 AGAAGGTTGCTATAGTGGATGCTCTCTGAGCTACACAAATTTCTCATGTAGAAAAGATG 2169  
Qy 1917 ACTTACCTTAAATGTGCGGCTCGAGTGTGTTTCCAGATGTGGGTAGAGG----- 1968  
Db 2170 ACTTACCTTAGACATGTGGACTGGAATTTTTTTTCATTAATGTGTTTCATCAAGTTTCACAT 2229  
Qy 1969 ---AGTTGACACAGGTTGATAGCATAAAGTCAGTCGCT 2003  
Db 2230 CTAAGTTGAAACAGGGTGTGTCTATAAAGTCAGTTTATCT 2267

RESULT 2

US-09-513-999C-3004  
; Sequence 3004, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; FILE REFERENCE: 59.US2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3004  
; LENGTH: 428  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS





GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5001  
; LENGTH: 5883  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-5001

Query Match 1.8%; Score 48.8; DB 3; Length 5883;  
Best Local Similarity 48.9%; Pred. No. 0.011;  
Matches 131; Conservative 0; Mismatches 137; Indels 0; Gaps 0;  
  
QY 342 AAGAGCGAAGAGAGAGATTAATATAGTCGCTGTTCTGNACTGTATGAACGAGCTGCTCG 401  
Db AGGTGGCCGACATGAATAAAGAGATGGAGACAGTGTGGGTGCTGGAAACTGCTGAGG 4153  
  
QY 402 TGCTTCAGCGGAAGAACCTGCTGGCCAGGAGCGGTGGAGACACAGAACTTTGAAGCTGG 461  
Db AGGTGAAGAGGAGCTCCAGAGGACCTGGAGGGCTGAGCCAGCGGCACGAGGAGAAGG 4213  
  
QY 462 CGAGTGACATGACACCACTGTCAGAGCTGCTACGCCAAACTTAAGGAGCAGTTGGAAACGT 521  
Db TGGCCGCTACGACAACTGGAGAGACCAAGACGCGGTGCGAGCAGGCTGGACGACC 4273  
  
QY 522 CCAGCGGGAGATGATCGGGCTTCAAGAGAGACAGGCGAGCTGAGTGCAGAACAGGA 581  
Db TGCTGGTGACCTGGACCAACCGCCAGAGCGCGTGCACCTGGAGAGAACGACGAAGA 4333  
  
QY 582 GTTTCATCAGCTCTCGAAGAAATGAGAA 609  
Db AGTTGACCAGCTCTCGCGGAGGAGAA 4361

RESULT 8  
US-09-949-016-13217/c  
; Sequence 13217, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13217  
; LENGTH: 28806  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13217

Query Match 1.8%; Score 48.8; DB 3; Length 28806;  
Best Local Similarity 47.4%; Pred. No. 0.03;

Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;  
  
QY 881 AAGCAGATCTCTGATCGAGAAACCGGAGCTGAAGAAGGTCTCTCCAGCAGATGAAGAAGGAG 940  
Db 25834 AAGGAGAGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAG 25775  
  
QY 941 ATGATCT 1000  
Db 25774 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25715  
  
QY 1001 GGCAGCTGTTCTATCTCCGATATAGAAGTACTCTCTGGGAACTCTGAGCAGAGACAGCCGTG 1060  
Db 25714 CAGACAGAGAGCGGAG 25655  
  
QY 1061 TGGGGCCCTTTCTCTGTACACTGTGTAGAGAGCAGCTGTGACAAACAGCATCAGGAAACAGTGG 1120  
Db 25654 AGAGAGAGACAGAGGGGTGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25595  
  
QY 1121 AGAATTTTGAAGTCACTGTAGAAAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAG 1180  
Db 25594 AGGATACATGAAATGATAGCAGATGAAATTCATACCCAGAGACCCAGGAGACACACAGAG 25535  
  
QY 1181 GGCCTTAA 1188  
Db 25534 AAACAGAA 25527

RESULT 9  
US-09-976-594-886  
; Sequence 886, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 886  
; LENGTH: 1312  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 981037.1  
US-09-976-594-886

Query Match 1.7%; Score 46.6; DB 3; Length 1312;  
Best Local Similarity 52.9%; Pred. No. 0.017;  
Matches 100; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
  
QY 376 TCTGAACCTGTATGAACCGAGCTGCTCGCTTTCAGCGGAAGAACTGTGTGGCCCGAGGAGAG 435  
Db 640 TCACTACTTTCTGCCCCAACCTCGACCTCTTTTCAGGGCAAGCCCAATTCGGCCCTGGAGAG 699  
  
QY 436 CGTGAGACACAGNACTTGAAGCTGGGAGTGACATGGACCACTGCAGAGCTCTAGCC 495  
Db 700 CGTGCCCAAGCAGACCTGGAGGTTGGCCAGGAAATTTCTACCAATCCCAAAAGCCTTGA 759  
  
QY 496 CAAACTTAAGGAGCAGTTGGAAACGTCAGCGGGAGATGATCGGGCTTCAAGAGAGAGA 555  
Db 760 CAACTATAGGCTCTGGGACTGCTTTGAAAGGACACAAACGGGGTGTCTCTCTCAGA 819  
  
QY 556 CAGCAGCT 564  
Db 820 CACAACT 828

RESULT 10





RESULT 14  
US-08-770-379-20/c  
; Sequence 20, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770.379  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 52342  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32207 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-770-379-20  
  
Query Match 1.7%; Score 46; DB 2; Length 32207;  
Best Local Similarity 42.7%; Pred. No. 0.21;  
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;  
  
Qy 393 AGCTGCTCGTCTCAGCGGAAGAACCTGCTGGCCCGAGGAGCGTGGAGACACAGAACT 452  
Db 20093 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 20034  
Qy 453 TGAAGCTGGGCGAGTGACATGACATGACATGACATGACATGACATGACATGACATGACAT 512  
Db 20033 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 19974  
Qy 513 TGGAAACCTCCAGGCGGAGATGATCGGGCTTCAAGAGAGAGACAGGCGAGCTGCAGTGCA 572  
Db 19973 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 19914  
Qy 573 AGACAGAGATTTGATCAGCTCTGGAAGATGAAAGATGAAAGATGAAAGATGAAAGATGAA 632  
Db 19913 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGGAGCAGCAGGATGAGCAGGAGCAGC 19854  
Qy 633 ATATCATAGCCAGCGGGCTACTCAGTATATCATGATGATGATGATGATGATGATGATGAT 692  
Db 19853 AGGATGACAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGC 19794  
Qy 693 ATAATAAGCTAAAGGAGCGCTTGCATCAGCTCGTTTATGAAACAAGAGGATAAAAACATAG 752

Db 19793 AGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAAACAGGAGC 19734  
Qy 753 CCATGATGTTTTAAATTATGTGGTGGAGCTGATGCCAACGAGGCTCATGGAGGACTG 812  
Db 19733 AGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGGAGGAGGATTTAGAGGAGCAGGAGCAGG 19674  
Qy 813 ACAAACAGAAAGCCAGGAATGAAGATGAGATGTACAAAATTCCTGTTGAATGATTATCAGT 872  
Db 19673 AGTTAGAGGATCAGGAGCAGGAGTTAGAGGAGCAGGAGGAGGATTTAGAGGAGCAGGAGC 19614  
Qy 873 ACCGCCAGAAAGCAGATCCTGATGGAGAAACCGGAGCTGGAAGAGGTCCTCCAGCAGATGA 932  
Db 19613 AGGAGTTAGAGGAGCAGGAGGAGTTAGAGGAGCAGGAGGAGGAGGATTTAGAGGAGCAGG 19554  
Qy 933 AGAAGAGAT 942  
Db 19553 AGCAGGAGTT 19544  
  
RESULT 15  
US-08-757-669A-20/c  
; Sequence 20, Application US/08757669A  
; Patent No. 6183751  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/757,669A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 45185-F  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 32207 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-757-669A-20  
  
Query Match 1.7%; Score 46; DB 3; Length 32207;  
Best Local Similarity 42.7%; Pred. No. 0.21;  
Matches 235; Conservative 0; Mismatches 315; Indels 0; Gaps 0;  
  
Qy 393 AGCTGCTCGTCTCAGCGGAAGAACCTGCTGGCCCGAGGAGCGTGGAGACACAGAACT 452  
Db 20093 AGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGC 20034  
Qy 453 TGAAGCTGGGCGAGTGACATGACATGACATGACATGACATGACATGACATGACATGACAT 512





**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:34 ; Search time 51 Seconds  
(without alignments)  
1055.516 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVDPVLTENKNLS.....CYGSSAFSAHGRDRLDLP 615

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/pCTUS\_COMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2152.5	68.0	504	2	US-10-104-047-3467
2	612.5	19.4	137	2	US-09-513-999C-7081
3	223.5	7.1	359	2	US-09-949-016-6507
4	222.5	7.0	905	2	US-09-248-796A-16333
5	222.5	7.0	2733	2	US-09-949-016-11433
6	220.5	7.0	1780	2	US-09-949-016-6899
7	220.5	7.0	1786	2	US-09-949-016-7880
8	215.5	6.8	2125	2	US-09-919-172-29
9	212.5	6.7	2863	2	US-09-538-092-1252
10	211.5	6.7	1388	1	US-08-685-576-1
11	211.5	6.7	2662	2	US-09-595-684B-31
12	210.5	6.7	1354	2	US-08-685-871-2
13	208.5	6.6	1388	1	US-08-685-576-4
14	208.5	6.6	1388	2	US-09-976-594-296
15	208	6.6	2704	2	US-09-538-092-1260
16	206.5	6.5	1676	2	US-09-949-016-7610
17	206.5	6.5	2186	2	US-09-949-016-10828
18	206.5	6.5	2349	2	US-09-538-092-914
19	204	6.4	1427	2	US-09-538-092-1044
20	203	6.4	3210	2	US-09-538-092-1154
21	203	6.4	3248	1	US-08-353-700-1
22	203	6.4	3248	5	PCR-US95-16216-1
23	200.5	6.3	1958	2	US-10-028-946-4
24	200.5	6.3	2053	2	US-09-964-956-11
25	200.5	6.3	2054	2	US-10-028-946-2
26	200.5	6.3	2066	2	US-09-964-956-9

ALIGNMENTS

RESULT 1

US-10-104-047-3467

; Sequence 3467, Application US/10104047

; Patent No. 6943241

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. 6943241el full length cdna

; FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25

; PRIOR APPLICATION NUMBER:

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 3467

; LENGTH: 504

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-104-047-3467

Query Match

Best Local Similarity 68.0%; Score 2152.5; DB 2; Length 504;

Mismatches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;

Qy	1	MGDWMVTDPVLTENKNLSQYTSKMSPSLSYQQVLCSSVPLSKNVHGVFGVCTGGE	60
Db	1	MGDWMVTDP-----ESKTSQYTSKMSPSLSYQQVLCSSVPLSKNVHSPFSAFCTED	56
Qy	61	NIEQISYLDQELTTFGFPSPSYEESKSKRELNIVAVLNCMELLVLRKNLLAQS	120
Db	57	NIEQISYLDQELTTFGFPSPSYEESKSKRELNIVAVLNCMELLVLRKNLLAQS	116
Qy	121	ETQNLKGSMDHLSQYAKLKEQLTSRRMIGIQERDRQIQCKNRSIHLQKNEKDE	180
Db	117	ETQNLKPSGMDHLSQYSLKKEQLTSRRMIGIQERDRQIQCKNRLHQLKNEKDE	176
Qy	181	OKLQNIASRATQYVNDVKRERENYKUKERHQLVMKKNKNIAMDVLNVYGRADG	240
Db	177	OKLQNIASRATQYVNDMKRERENYKUKERHQLVMKKNKNIAMDVLNVYGRADG	236
Qy	241	SWRTGKTARNEDEMYKILLNDYEROKQILMENAELKKVLOOMKEMISLLSP	300
Db	237	SWRTGKTARNEDEMYKILLNDYEROKQILMENAELKKVLOOMKEMISLLSP	296
Qy	301	EAEDEGTGTAISDIEDSGELSDVSWGLSCDTVREQLTNSIRKQWILKSHVEK	360
Db	297	ERVDDSTGTV-ISDVEEDAGELSRWSMDLSCETVREQLTNSIRKQWILKSHVE	355
Qy	361	ASKVHSEGLNEEDVISRQDHEQTEKLETERCKEMIKAAQOQLLQOQLATTCD	420

Db 356 VSKVHLEGENDEVISRQHEQETEKLELEIQCKEMIKTQQLQQLATAYDDTTSL 415  
Qy 421 LRDCYLLEEKERLKEWTLFKQCKNFERERSFTEAAIRLGL 464  
Db 416 LRDCYLLEEKERLKEWTLFKQCKNFERERSFTEAAIRLGL 459

RESULT 2  
US-09-513-999C-7081  
; Sequence 7081, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59.U2.REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7081  
; LENGTH: 137  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-513-999C-7081

Query Match 19.4%; Score 612.5; DB 2; Length 137;  
Best Local Similarity 87.0%; Pred. No. 1.8e-39;  
Matches 120; Conservative 10; Mismatches 7; Indels 1; Gaps 1;  
Qy 272 MENAELKKVLOQMKEMISLLSPQKKKPRERAEDGTGTVASIDIEDDGGELSRDWSGLS 331  
Db 1 MENAELKKVLOQMKEMISLLSPQKKKPRERVDSTGTV-ISDVEEDAGELSRDSMDLS 59  
Qy 332 CDTVREQLTNSRKQWRILKSHVEKLDNQASKVHSEGLNEEDVISRODHEQETEKLELEI 391  
Db 60 CDTVREQLTNSRKQWRILKSHVEKLDNQASKVHLEGFENDEVDISRQHEQETEKLELEI 119  
Qy 392 ERCKEMIKAQOQLQOOL 409  
Db 120 QCKEMIKTQQLQOOL 137

RESULT 3  
US-09-949-016-6507  
; Sequence 6507, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6507  
; LENGTH: 3259  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6507

Query Match 7.1%; Score 223.5; DB 2; Length 3259;  
Best Local Similarity 19.9%; Pred. No. 1e-07;  
Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;  
Qy 15 ENKLSQVTSKMS-----PSSLYSQVLCSSVP-LSKNVHGVGVPFCTGENTIRQ 64  
Db 1225 ENENIGDQLRQLQVRESIDGKLPS-DOQESCSSTPGLEELPKATEQHHTQPVLES 1282  
Qy 65 SI-----SYLDOELTTFGFPSIYE---ESKSEKARELNIVAVLNCMELLVLQRKNLLA 116  
Db 1283 NLCPDWPSPSHSDASALQGTVAQIKAQLEIAEKVLELKVSTTSBELTKKSEVQOL 1342  
Qy 117 QESVETQNLKLS--DMDHLQSCYAK-LKEQLETSRRREMIGLQERDRDQLCKNRSLSHQLL 173  
Db 1343 QEIQNKQGLETESLKTVSHEAEVHAESLQCKLESSQLQIAGL-EHLRELQPK----- 1393  
Qy 174 KNEKDEVQKLQNIITASRATQYNHDVKERYNKLKERLHQLVMNKKDKNTAMDVLNVYG 233  
Db 1394 ---LDELQKL-----ISKKEEDVSVLSGQL-----SEKEAALTKI----- 1425  
Qy 234 RADGKRGSWRTDKTEARNEDEMYKIL-----LNDVEYRQKQILMENAELK----- 278  
Db 1426 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLQVELCEHKKQKPEETGE 1473  
Qy 279 --KVLQMKKEMISLLSPQK-----KKPRERAEDGTGTV-----AISDIED--DSGELS 323  
Db 1474 ESRKQIQIRKLAALISRKALKENKSLQELSARGTIERLTKSLADVESQVSAQNKE 1533  
Qy 324 RDSVWG-----LSCDTR-----FQL 339  
Db 1534 KDTVLGRLLALQEBRDKLITEMDRSLLENQSLSSCSLKLAEGLTDEKELVKEIESL 1593  
Qy 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVISR 377  
Db 1594 KSSKIARSTEQEKGKELQKSEYIELLSQSYENVNSHAEIRIQHVAVROEKQELVCKLRST 1653  
Qy 378 QDHEQETEK-----LELEIERCKEMIK-----AQOQLQOQ-----LATTC 413  
Db 1654 EANKKETEKQLQEAQEQEMEEKMKRPFKSKQKQKILEEBENDRLRAEVHPAGDTAKEC 1713  
Qy 414 DDDTSLLRDYLLEEKERLKEWTLFKQCKNFERERSFTEAAIRLGLERKAFEEBRA 473  
Db 1714 ME--TLSSNASMKEELERVMEYETLSKFKQSLMSEKOSUSEVEVDL-----KHQIEDNV 1767  
Qy 474 SWVKQOFLNMTNFDHNSENVK---LFSAFSGSDPDNLI VHSRPRQKHLHSVANGVPAC 530  
Db 1768 S--KQANLEATE-KHDQNTVTEEGTQSI PGETESEQDSLSMSTR-----PTC 1811  
Qy 531 TSKLTKSLPASPSTS-DPROTHSCVSBHSSIS-----VLNITPBESKPSSEVAR 577  
Db 1812 SESVPSAKSANPAVSKDF-----SSHDEINNYLQOIDQLKERIAGLEEBEKQKNKPSQ 1864  
Qy 578 ESTQCKWSVQSRPSSREG 595  
Db 1865 TLENEKNTLSQISTKDG 1882  
RESULT 4  
US-09-248-796A-16333  
; Sequence 16333, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keitch Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208

```
; SEQ ID NO 16333
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-16333

Query Match
 7.0%; Score 222.5; DB 2; Length 905;
Best Local Similarity 18.1%; Pred. No. 2e-08;
Matches 137; Conservative 140; Mismatches 264; Indels 217; Gaps 27;

QY 10 PVLCTENKNLSQYTSMTKSPSSLYSQVLCSSVPLSKNVHGVFGV-FCVTGEN-----IE 63
Db 143 PIL-----LDAYKIEKKWEIIRIPKVLQENTPVSSSEGLVLYNGLNKEKTKLN 196

QY 64 QTSISYLDQELTTGFPSPLYEESKEAKRELNIVAVLNCMNEILLVLR-KNLLAQ----- 117
Db 197 GEIRKLTTKDNNTWFKIV-----KLEESNKLN-----NDVKRLDTMKEKEVNGKKIDS 247

QY 118 ---ESVETONKLGSDMDHLQCYAKLKEQLETSRREMIGLQERDRQCKNRSLSHQLL 173
Db 248 SSEELEINQNDKLTLDKDK-----EKIESLKLEITDRDNKTKDLESNTSSNSEL 300

QY 174 KNEKDEVQKL-----QNIASRATQVNHVDVKERE 204
Db 301 NTELEPKEVPTKVNQDKSSSGEENTDKDSSVGDSSDNDSDVLTNLSQLSKSLSTKETE 360

QY 205 YNKLKERLHQLVMNKDKN-----IAMDVLNYYGRADGKRGSWRTDKTEARNEDEMY 256
Db 361 VEELTNEVRTLSQLDNKNEEIEDRLDLVKEIGNELVTSKDEIKSLKNSQKSDNEDSVT 420

QY 257 K-----ILLADYEVROQIILMENAELKKVLQOMKKEMISLLSPKKKPRERAED----- 305
Db 421 KEDTNTQINDWELVNSQLKENKELKEELDVATKE-----SKEKEKERRKSEDLQRKQIQ 475

QY 306 -----GTGTVAISDIEDDSGEL-----SRDSVWGLSCDTVRBQLTN 341
Db 476 TLQKLESTENKESKTKDLKNLLEKEKELEKREISLSKFNKSDSLKLEISLSKSLTN 535

QY 342 -----SIRKQWRILKSHVEKLDNQASKVHSEGL-----NEEDVIRSQD----- 379
Db 536 KONSINELKTQVDELQOSKNSLKVSELSKSNELQSNMDFLKDKNELLTKQEVLMEN 595

QY 380 -----HEQETEKLELETERCKE-----MIKAQOQ 403
Db 596 TKSLLNQVTKLOQKQDVITELEKTKNKLDIVIADKSQANDMLSYKKQHEELMMKSKY 655

QY 404 LLQOQLATTCDDDTTSLLRDCVYLLEKER-----LKEEWTLFKEQKNFERERRSFT 455
Db 656 SLRIE---SLEDDLTEARN---LQERTRETSNMRLLVDABEMLKQOQDSKLE----- 704

QY 456 EAAIRLGLERKA-FEERASWVKQFLNMTNFDH-----QNSENVKLFSAFGSSD 505
Db 705 ---IARALBDKABIERNANSLIRKQRELDKESVQEVKLVKELEN-KVMTLEKEEKS 760

QY 506 PDLNLIHSPROKHLHSHVANGVAPACTSKLTSLPASPSTSDPROTHSCVSEHSSI----- 560
Db 761 KESLSDNTQMSK---ELSSSTIETLRTALNNSTNKIRLDENYNNNNKLSEDTNRTLRFEL 818

QY 561 ----SVLNTPTBESKEPEVAREBSTDQKWSVQSRPSRE 594
Db 819 SKNYKILNQOQYQMKERKGSSESTKQDKPVEAPENSND 856

RESULT 5
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-11433

Query Match
 7.0%; Score 222.5; DB 2; Length 2733;
Best Local Similarity 18.9%; Pred. No. 9.5e-08;
Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;

QY 15 ENKNLSQYTSMTKMS-----PSSLYSQVLCSSVP-LSKNVHGVFGVFCVTGENIEQ 64
Db 699 ENENIGDQLRQLQIQVRESIDGKLPTST--DQESCSSTPGLSEPLFKATEQHHTQPVLES 756

QY 65 SI-----SYLDQELTTGFPSPLYE-----ESKSKEAKRELNIVAVLNCMNEILLVLRKNLLA 116
Db 757 NLCPDWPSPHSDASALQGGTSVAQIKAQKLEIAEKVELELVKVSSTTSELTKKSEEVFQL 816

QY 117 QBSVETQNLKLAGS--DMDHLQSCYAK-LKEQLETSRREMIGLQERDRQCKNRSLSHQLL 173
Db 817 QBOINKGLEIESLTKTVSHEAEVHAESLQCKLESSOLOIAGL-EHURELOPK----- 867

QY 174 KNEKDEVQKLQNIIASRATQVNHVDVKEREYNKLERLHQLVMNKDKNIIAMDVLNYYVG 233
Db 868 ---LDELQKL-----ISKKEEDVSYSGLQ-----SEKEAALTKI----- 899

QY 234 RADGKRGSWRTDKTEARNEDEMYKIL-----LNDYEVROQIILMENAELK----- 278
Db 900 -----OTEIIEQEDLIALHTOLEWQAKEHDERIKQLQVELCEMKQKPEEIGE 947

QY 279 ---KVLQOMKKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323
Db 948 ESRAKQIQORLKQALALSRKEALKENKSLQELSLARGTIERLTKSLADVESQVSAONKE 1007

QY 324 RDSVWG-----LSCDTVVR-----EQL 339
Db 1008 KDTVLGRLLALQERDKLITEMDRSLLENQSLSSCSLKLALGLTEDEKLVKEIESL 1067

QY 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVISR 377
Db 1068 KSSKIAESTEWQEKHKEQKEYEILLQSYENVSNBAERIQHVVEAVRQEKQBLYGKLRST 1127

QY 378 QDHEQETEK---LELEIERCKEMIK-----AQOQLLOOQ-----LATTC 413
Db 1128 EANKKETEQLOEAQOEEMKEKMKRFAKSKQKILEEENDRLRAEVHPAGDTAKEC 1187

QY 414 DDDTTSLLRDCVYLLEKERLKEEWTLFKEQKNFERERRSFTAAIRLGLERKAFEEPERA 473
Db 1188 ME--TLLSSNASKMEELERVKMEYETLSKKFOSLSEKDSLSEEVQDLKHQIEGNVSKQA 1245

QY 474 SW-VKQOFLNMTNFDHNSENVKLFSAFGSSDDPNLIHVSPPROKHLHSHVANGVAPACTS 532
Db 1246 NLEATEKHNDQNTNVTGQTQISIP-----GETEEQDSLMSMSTR-----PTCSE 1287

QY 533 KLTKSLPASPSTS-DFROTHSCVSEHSSIS-----VLNITPEESKPSPEVARES 579
Db 1288 SVPSAKSANPAVSKDF-----SSHDEINNYLQIIDQIKERIAGLEBEKQKNKEFSQTL 1340

QY 580 TDQKWSVQSRPSREG 595
Db 1341 ENKNTLLSQISTKDG 1356

RESULT 6
```

US-09-949-016-6899  
; Sequence 6899, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6899  
; LENGTH: 1780  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6899

Query Match 7.0%; Score 220.5; DB 2; Length 1780;  
Best Local Similarity 20.6%; Pred. No. 7.4e-08;  
Matches 156; Conservative 117; Mismatches 235; Indels 251; Gaps 36;

Qy 16 NKNLSQVSTKMSPSLSY-----QQVLCSSVPLSKNVHGVGVFCTGENIEQSI 66  
Db 996 SKQVKYRIQEPNRENSPHSSIEAIWEECKEIVKAS---SKSKHQI-----EELEQOI 1045  
Qy 67 SYLDQELTTGFPFSLYEESKREKELNIVAVLNCMNEILLVLRKNNLLAQS-----VET 122  
Db 1046 EKLAQEVK--GYKDENNRLKEKHKNQDILLK-----EKETLIQQLKEELQKNVTLDVQI 1099  
Qy 123 QNLKLG-----SDMDHLQSCY-AKLKEQ-----LETSR-----REMIGLQERD 159  
Db 1100 QHVVEGKRALSSELTQGVTCYAKIKELETTILETKQVRSKSHSAKLQDILEKESI-ILKLE 1158  
Qy 160 RLQCKNRSIHLQKNEKD-----QQVLCSSVPLSKNVHGVGVFCTGENIEQSI 181  
Db 1159 RNLKEFQEHQDSVKNKTDNLNVKELKKEEITQLTNNLQDMKHLQLKKEEBEETNRQETE 1218  
Qy 182 KL-QNTIASRATQYN--HDVKKEREYNKLERHLQVNMKKDKNIAMDVLNVYGRADGK 238  
Db 1219 KLKEELSASSARTONLKAQDRKEEDYADLKEKLT--AKKQIKQVQKEV-----1266  
Qy 239 RGSWRDTKEARNEDEMYKILLNDYEQKQILMENAELKKVLOQMCKEMISLLSPQKK 298  
Db 1267 -----SVMRDEKLLRIKINELEKKKQCSQELDMKQRTIQQLEQLNNQKVEEAIQ 1318  
Qy 299 PRERA-----EDGTGTVASIDEDSGELSRSSVWGLSCDTVREQLTNSIRKQW 347  
Db 1319 QYERACKDLNVKEKIIDMRMTLE-----EQEQTQVEQDQVLEAKLEEV-ERLATELEK-W 1372  
Qy 348 RILKSHVEKLDNQAQSVHSEGLNEEDVISR-----QDHEQETEKLEIERCKEMIKAAQQ 403  
Db 1373 KEKCNDETKNQRSNKEHE--NNTDVLGKLTNLQDELQESQ-KYNADR-KKMLEEKKM 1428  
Qy 404 LLOQQLATTCDDDTTSLLRDCLLBEKERLKEEWTLFKEQKNKFRERRSF-----454  
Db 1429 LITQ-----AKEAENIRN-----KEMKKAEDRRFPFQKNEMEIL 1464  
Qy 455 TEAAIRGLGERKAFEBEER-----ASWVKQFLNMNFDQN-----SENKLFSAF 500  
Db 1465 TAOLTEKSDLOKWEERDQLVAALIEQLKALISSNVQKDNIEQLKRIISSETSIETQI 1524  
Qy 501 S-----GSSDPDNLVHSRPRQKKLHVSANGV-----527  
Db 1525 MDIKPKRISADPKL--QTEPLSTSFETSRNKIEDSGSVLDSCEVSTENDQSTFRFPKE 1582  
Qy 528 -----PACTSKLTKSLPASPSSTDFROTHSCVSEHSSI---SVLNITP 567

```
QY 455 TEAAIRGLERKAFEEB-----ASWVKQOFLNMTNFDHON-----SENKLFSAF 500
Db 1471 TAQTEKSDLOKWRERDQVLALEIQLKALISSNOKONEIEQLKRIISSETSIETQI 1530
QY 501 S-----GSSDPNLIHVRPRQKKLHVSANGV-----527
Db 1531 MDIKPKRISSADPKL--QTEPLSTSPFISRNKIEDGSVVLDSCEVSTENDQSTRPKPE 1588
QY 528 -----PACTSKLTKSLPASPSTSDPROTHSCVSEHSSI-----SVLNITP 567
Db 1589 LETQFTPLQPNKMAVKGPGCTTPVTVKIPKA-----RKRKSNEMBEDLVKCNKKNATP 1642
QY 568 EESKP---SEVARESTDOKSVOSRPSREGCYSGCSSA 603
Db 1643 RTNLKFPISDRNSSVVKQKVAIRPSSKK-TYSLRSQA 1680

RESULT 8
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. 6673545
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673545 3774181CD1
US-09-919-172-29

Query Match 6.8%; Score 215.5; DB 2; Length 2125;
Best Local Similarity 18.9%; Pred. No. 2.3e-07;
Matches 142; Conservative 136; Mismatches 260; Indels 213; Gaps 31;

QY 1 MGDWMTVDPLCTENKILSQYSETKMGSPSSLYSQOVLCSVPKSNVHGVFGVCTGE 60
Db 452 LDDWI---QOVETTKIOENQENSKTLATQNLQOKMLVSEIEMKQSKWDE---CQKY 504
QY 61 NIEQISYLDQELTTGFPSPLYEESKSEAKR-----ELNIVAVLNCM 103
Db 505 AEQYSATVKDYELQNTYRAMVDSQOKSPVKRRRMQSSADLIQEFMDLRTTRYALTVM 564
QY 104 NELL-----VLQKNLLAQESVETONKLGSDMDHL 134
Db 565 TQYIKFAGSLKRLBEEETKRCETSEHGAYSDLLQ-----QKATVLENSKLTGKISEL 619
QY 135 QSCYAKLKEQLETSRREMIGLOE-RDRQLQCKNRSLSHLQKNEKDEVOQLQNIIA-SRAT 192
Db 620 ERWAEKQKQSRVEBELPKVREAAENELRKQQRNVEDI-----SLQIKRAESEAK 670
QY 193 QYNHD-----VKRERYNKLKERLHQLVMNKKDKNIAMD--VLNVYGRADGKRGSRWTDK 246
Db 671 QYRRELETTVREKEAAREL-ERVROLTIEAEKRAAveenLLNFRNQLEENTFTFRRTLE 729
QY 247 TEARNEDEMYILLNDYVROKQILME-----NAELKKVLQOMKKEMI-----289
Db 730 DHLKRD-----LSLNDLE--QQRNKLMEELRRKRDNEEBELKLIKQEKDLAFQKQVAKQ 784
QY 290 -----SLSPQKKKPREAED-----305
Db 785 LKEKOKIELEARRKITEIQYTCRENALPVCPIQTQATSCRAVTGLQOEHDQKQAEELKQOV 844
```

```
QY 306 GTGTVAISDIEDSGELSRDSVWGLSCDVTREOLTNISIRKQWRILKSHVEKLN-----359
Db 845 DELTAANRRKAEQDMRELTYE-----LNALQLEKTSSEKA-RLKDKLDETNTLRCLK 897
QY 360 -----QASKVHSEGLNE-----EDVIS-----RODHEQETEKLELE--390
Db 898 LELERKQAEKGYSOQLRELQNLQNTTKAEEMQEAADLKKIKRNYQJLESLSNHEKG 957
QY 391 -----IERKEMIKAQOQLLQOQLATTCCDDDTTSLLRDCYLLBEEKERLKEEWTLFKE 442
Db 958 KLQREVDRITRAHAVAENKIQHLNSQIHSPRDEKELERLQICQ--RKSDHLKEQFEKSHE 1015
QY 443 Q-KKNFERERRSFTEAAIRGLERKAFEEERASWVQOFLNMTNFDHONSENKLFSAFS 501
Db 1016 QLLQNIKAEBEN-NDKIQRLNEELEK-SNECAEMLKQKVEELT---RONNETKLMQRIQ 1070
QY 502 GSSDPNLIHVRPRQKKLHVS---ANGVPACTSKLTKSLPASPST-SDPROTHSCVSEH 557
Db 1071 AES--ENIVLEKQTIQORCEALKIQADGFKDQLRSTNEHLHKKOTKTEQDFQRIKCLDE-1127
QY 558 SSISVLNITPEESKPSSEVARESTDQKWSVQS 588
Db 1128 -----DLAKSQNLVSEFKQKCDQQNIION 1152

RESULT 9
US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 6.7%; Score 212.5; DB 2; Length 2663;
Best Local Similarity 22.4%; Pred. No. 5.4e-07;
Matches 141; Conservative 114; Mismatches 225; Indels 149; Gaps 30;

QY 17 KNLQYTSSTKSPSSLYSQOVLCSVPKSNVHGVFGVCTGENIEQISYLDQELTTF 76
Db 971 ESLKQH-QETINTLASKISEEV-----SRNLHMEENTGETKDFQKQKMGIDKK----1018
QY 77 GPSPLYESKSEAKRELNIIVAVLNCMNEELLVLQKNL-LAOSVETQNLKLGSDMDHLQ 135
Db 1019 -----QDLKAKNTQTLTADVK-DNEIIBQQRKIFSLIQEKNELOQM-----LE 1060
QY 136 SCYAKLKEQLETSRREMIGLOERDQLOCKNRSLSHLQKNEKDEVOQLQNIASRATQVN 195
Db 1061 SVIAB-KEQLKTDLKENI-----EMTINQEELURLL---GDELKKQOEIVAQ---EKN 1106
QY 196 HDVKKEREYNKLKERLHQLVMNKKDKNIAM-----DVLNVYGRADGKRGSRWTDK 250
Db 1107 HAIK-KEGELSRTCDRLAEVEEKLKESQOQKQQLNV-----QEESEMQ 1154
QY 251 ---NEDENMYILLNDYVROKQILMENAEELKKVLQOMKKEMISLLSPQKKPREAEDGT 307
```





```
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
; US-09-595-684B-31

Query Match          6.7%; Score 211.5; DB 2; Length 2662;
Best Local Similarity 22.1%; Pred. No. 6.4e-07;
Matches 141; Conservative 116; Mismatches 233; Indels 147; Gaps 30;

QY 14 TENKLSQYTSKMPSSLYSQVLCS-----SVPLSKNVHGVFGVCTGENIEQSISY 68
Db 955 TVNNWIDT-QEQLRNALESKHQETINTLKSKISEEVRNLHMEENTGETKDFQKQWVG 1013
QY 69 LDELITFFGPPSYIESKSEAKRELNIIVAVLNCMELLVLQKNI-LAQSVEVTONLKL 127
Db 1014 IDKK-----QDLKANTQTLTADV-K-DNEIIBQQRKIFSLIQEKNELQOM-- 1057
QY 128 GSDMDHLQCYAKLKEQTSREMIQLQDRDQLQCKNRSLLHLLKNEKDEVQKLNII 187
Db 1058 -----LESVIAE-KEQLKTDLKENI-----EMTIENQEEURLL---GDELKQOEIV 1100
QY 188 ASRATQYNDHVKRREYNKRLKRLHQLVMNKKDKNIAM-----DVLNVYVGRADGKRGSW 242
Db 1101 AQ---EKNHAIK-KEGELSFTCDRLAEVBEKLEKSQLQOEKQOQLNV----- 1145
QY 243 RTDKTEAR---NEDEMYKILLNDYEQROKILMENAEKLVLOOMKKEMISLLSPQKKP 299
Db 1146 QEEMSEMOKKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERK--- 1202
QY 300 RERAEDGTGVAISDIEDDSGELSRDSVNGLSCTDVTREQLTNSIRKQWRI-----LKSH 353
Db 1203 -----VLKELQ-KSFETERDLRGY-----IREIATGLQTKBELKIAHHLKHEH 1246
QY 354 VEKLDNQASKVHSEGLNEEDVISRDHQBTEKLELEI-----ERCKEMIKAQOQ 403
Db 1247 QETIDELRSVSEK---TAQINTQDLEKSHYKLOEEI PVLHHEEQELLNVKKVSTQET 1303
QY 404 LIQOQLAT--TDDDTTSLRLDCYLLEEKERLKEEWTLFKQKKNFERRRSFTAAIRL 461
Db 1304 MNELELLTQSTTKDSTTLAR---IEMERLRLNEKFOESQOEIKSUTKERDNL--KTIKE 1358
QY 462 GLE-----RKAFESERASWKO-QFLANMTNFDHONSENKVLFSAF----- 500
Db 1359 ALEVHDQLKEHIRETLAKIQESQSQKQSLNNKEKNEDTTKIVSMEQFKPKDSALLRI 1418
QY 501 -----SGSSDPDNLIVHSRPRQKKLHSHVANGVPACTSKLTKSLPAS--PSTSDF 547
Db 1419 EIEMLGLSKRLQESHDEMSKVAKEKDDLQRLQEVLSQESDQLKENIKETIVAKHLETEBEL 1478
QY 548 ROTHSCVSE-HSISVNLNTPESKPKSEVARESTDQK 583
Db 1479 KVAHCLKEQEETINELRNLSE-KETEI---STIQK 1511
```

RESULT 12

```
US-08-685-871-2
; Sequence 2, Application US/08685871
; Patent No. 6013499
; GENERAL INFORMATION:
; APPLICANT: NARUMIYA, Shuh
; APPLICANT: INAMATSU, Akihito
; TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,871
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184102
; FILING DATE: 25-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-262553
; FILING DATE: 14-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/845
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-685-871-2

Query Match          6.7%; Score 210.5; DB 2; Length 1354;
Best Local Similarity 21.7%; Pred. No. 3e-07;
Matches 134; Conservative 103; Mismatches 219; Indels 161; Gaps 26;

QY 83 EBSKSEAKRELNIIVAVLNCMELLVLQKKNLLAQESVETQNLKSGSDMDHLQSCYAKLK 142
Db 570 ENSKS-----ISQLESNLR--ELQERNRI-----LENSKQSTDKD-----YYQLQ 607
QY 143 EOLETSSRR-----EMIGLOERDROLQCKNRSLLHLLKNEKDEVQKLNIIASRATQYNNH 196
Db 608 AILEAERRDRGRHSEWIG-----DLQARITSLQEEVVKHLKHNLEKVEGERKEAQDMLNH 661
QY 197 DVYRKER-----EYNKLERLHQLVMNKK-----DKNIAMDVLNVVG--RADGKRG 240
Db 662 SEKEKNNLIEDLNYKLKSLQORLEQEVNEHKVTKARLTDKHQSIEEAKSVAMCEMKK 721
QY 241 SWRTDKTEARN-----EDMYKILLNDYEQR 267
Db 722 EBERAREAKENRVVQIEKQCSMLDVLQKSOQKLEHLTGKNERMEDEVKNLTLQEQESN 781
QY 268 KOILMENA-----ELKKVLOQMCKEMISLLSPQKKKPRERABDGTGTVAISDIEDD 318
Db 782 KRLLLQNLKTOAFEDADNLKGLKQKQKQEQINTLLEAKRLLLEFELAQ-----LTKQYRG 835
QY 319 SGELSRDSVNGLSCTDVTREQLTNSIRK-QWRILKSHVEKLDNQASKVHSEGLNEEDVISR 377
Db 836 EGQMR-----LQDQLEAEQVFTLYKTQVKELKEEIEKNRENKLTQELQONKETLAT 890
QY 378 Q-----DHEQETEKL-----ELEIERCKEMIKAQOQLLQOQLATTCDDDTTSL--R 422
Db 891 OLDLAETKAESQELARGLEEQYFELTQESKKAASRNQEIITDKOHTVSRLEANSMLTK 950
QY 423 DCYLL-----EKEERLKEEWTLFKQO-----KKNFERRRSFTAAIRIG--L 463
Db 951 DIEILRRENEELTEKMKKAEYKLEKEEISNLKAAFEKNINTERTLTKTQAVNKLAEIM 1010
QY 464 ERKAFEEERASWVKQFLNMTNFDHONSENKVLFSAFSGSSDPDNLIVHSRPRQKKLHSHV 523
Db 1011 NRKDFKIDRK-----ANTQDLRKKEKENKRLQLELNQERKFNQWVVK--HQELNDM 1062
QY 524 -ANGVPACTSKLTKSLPASPTSDPROTHSCVSEHSSISVNLNTPESKPKSEVARESTDQ 582
Db 1063 QALVEECAHRLNQLQASKESDIEQLRA-----KLIDLSDSTSVASFPSADETD- 1113
```

```
Qy 583 KWSVQSPSSR-EGCYS 598
Db 1114 ----GNLPESRIEGWLS 1126

RESULT 13
US-08-685-576-4
; Sequence 4, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akihiro
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819uaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-325129
; FILING DATE: 20-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-17150
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-131206
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16887/843
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-685-576-4

Query Match 6.6%; Score 208.5; DB 1; Length 1388;
Best Local Similarity 20.1%; Pred. No. 4.4e-07;
Matches 109; Conservative 109; Mismatches 186; Indels 139; Gaps 21;

Qy 67 SYLDDELTFGPPSLYVESKSKAKRELINAVLNCMNEILLVLRKLLAQAESVETQNLK 126
Db 402 AFVGNQLPFIQF-TYREN-----LLSDSPSCRENDISQSRKN---RESQBIQK-K 448

Qy 127 LGSMDHDLQSCYAKLKEQLTSRRREM-IGLQERDRQLQ---CKNRSLLHQLLNKNEDEVOK 182
Db 449 LYTLEBHL-SNEMQAKELEQCKSVNTRLEKTAKELEBEEITLRSVESALRQLERKAL 507

Qy 183 LQNTIATSRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 508 LQHKNAEYORKADHEADKKRNLENDVNSLKQLEDLKKRNQNSQISTEKNVQLORQOLD 567

Qy 193 LQNTIATSRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 508 LQHKNAEYORKADHEADKKRNLENDVNSLKQLEDLKKRNQNSQISTEKNVQLORQOLD 567
```

```
Qy 239 RGSWRD-----KTEARNEDEMYKILLNDYEVROKQIILMENAELKKVLQOMKKEMIS 290
Db 568 NALLRTESDTAARLRKTAQESSKQIQOLESNNRDLQDKNCLETTAKLK-----LEKEFIN 622

Qy 291 LLSPOKKKPRERA-----EDGTGTVA-----ISDIEDD 318
Db 623 LQSALESERRDRTHGSEIINDLQGRICGLBEDLKNGKILLAKVELEKRLQLOERFTDLKE 682

Qy 319 SGEISRDSVWGLSC-----DTVREQLTNSIRK-----Q 346
Db 683 KSNMEIDMTYQKLVQIQOSLEQSEAEHAKTKARLADKNKIYSEIEAKSEMKEKKLLE 742

Qy 347 WRILKSHVEKLDNOASKVHS-----EGLNEEDVISQDH-----EQETEK 386
Db 743 ERTLKQKVENLLLEAEKRCSLDDCDLKOSQOKINELLKQKDVNLNEDVRNLTLIKIEQTK 802

Qy 387 -----LELEIERCKEMIKAQOQLLQ-----QLATTCDDDTTSLLR-----DCYLLEE 429
Db 803 RCLTQNDLKMOTQVNTLKMSEKQLKQENNHLMEMKNLKNQNAELRKERQDADQOMKEL 862

Qy 430 KERLKEW---TLFEKQCKNFERERRSFTEAAIRLGLERKAFEEERASWVQOFLNMTNF 486
Db 863 QDQLEAEQYFTLYKTQVRELKESCEKTKLGLKELQOQKQELQDERDSLAAQLLEITLTKA 922

Qy 487 DHQ 489
Db 923 DSE 925

RESULT 14
US-09-976-594-296
; Sequence 296, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 296
; LENGTH: 1388
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 064987CD1
US-09-976-594-296

Query Match 6.6%; Score 208.5; DB 2; Length 1388;
Best Local Similarity 20.1%; Pred. No. 4.4e-07;
Matches 109; Conservative 109; Mismatches 186; Indels 139; Gaps 21;

Qy 67 SYLDDELTFGPPSLYVESKSKAKRELINAVLNCMNEILLVLRKLLAQAESVETQNLK 126
Db 402 AFVGNQLPFIQF-TYREN-----LLSDSPSCRENDISQSRKN---RESQBIQK-K 448

Qy 127 LGSMDHDLQSCYAKLKEQLTSRRREM-IGLQERDRQLQ---CKNRSLLHQLLNKNEDEVOK 182
Db 449 LYTLEBHL-SNEMQAKELEQCKSVNTRLEKTAKELEBEEITLRSVESALRQLERKAL 507

Qy 183 LQNTIATSRATQYNHDKVRK---EREYNKLERLHOLVNNKKDKNTAMDVLNVVGR-ADGK 238
Db 508 LQHKNAEYORKADHEADKKRNLENDVNSLKQLEDLKKRNQNSQISTEKNVQLORQOLD 567

Qy 239 RGSWRD-----KTEARNEDEMYKILLNDYEVROKQIILMENAELKKVLQOMKKEMIS 290
Db 568 NALLRTESDTAARLRKTAQESSKQIQOLESNNRDLQDKNCLETTAKLK-----LEKEFIN 622
```

```
QY 291 LLSPOKKKPERA-----EDGTGTVA-----ISDIEDD 318
Db 623 LQALSERRRDRTHGSEIINDLOGRIGUEEDLKNGKILAKVELEKRLQERFTDLEKE 682
QY 319 SGELSRDSVWGLSC-----DTVREQLTNSIRK-----Q 346
Db 683 KSNMEIDMTVOLKVIQOSLEQEBEAHKATKARLADKKNKIYESIEEAKSEAMKEMEKLLJE 742
QY 347 WRILKSHVEKLDNQASKVHS-----EGLNEEDVTSRODH-----EQETEK 386
Db 743 ERTLKQKVENLLLEAEKRCSLDCDLKQSOOKINELLKQKVLNEDVRNLTLLKIEQETOK 802
QY 387 -----LELEIERCKEMIKAAQOQLLOQ-----QLATTCDDDTTSLLR-----DCVLLLE 429
Db 803 CCUTQNDLKMQTQOVNTLKNSEKQKQOENNHLMEMKQNELEKQNAELRKERQDADGOMKEL 862
QY 430 KERLKEW---TLFKEQKKNFERRSFTEAAIRLGLERKAFEEERASVWVKQOFLMNTNF 486
Db 863 QDOLEAEQYFSTLYKTQVRELKECECKTYLKGELQOQKQELQDERDLSLAQLEITLTKA 922
QY 487 DHQ 489
Db 923 DSE 925
```

## RESULT 15

```
US-09-538-092-1260
; Sequence 1260, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1260
; LENGTH: 2704
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q03001
US-09-538-092-1260
```

```
Query Match 6.6%; Score 208; DB 2; Length 2704;
Best Local Similarity 18.5%; Pred. No. 1.2e-06;
Matches 142; Conservative 136; Mismatches 260; Indels 228; Gaps 31;
```

```
QY 1 MGDWMTVPDLCTENKNLSQYTSFKMSPSLYSQOVLCSVPKSNVGHVGVGFCTGE 60
Db 1016 LDDWI---QVETTKIQENQSPNSKTLATQNLQKMLVSEIMKQSKWDE---CQKY 1068
QY 61 NIEQISYLDQLTTFGPPSLYEESKKEAKR-----ELNIVAVLNCM 103
Db 1069 AEQYSATVKDYELQMTMYRAMVDSQKSPVKRRMOSADLIIBFMDLTRYTALTILM 1128
QY 104 NELL-----VLOEKNLAAQES 119
Db 1129 TOYIKFAGDSLKRLEEEVINSHPVNIKELEIKRCKTSEHGAYSDLLQR-----QKA 1183
QY 120 VETQNLKGSMDHDLQSCYAKLKEQLTSRRRMIQLQE--RDRLOQCKNRSLSHLKNEKD 178
Db 1184 TVLENSKLTGKISELERMVAELKQKSRVEELPKVREAENELRKQQRNVEDI----- 1237
```

```
QY 179 EVQKLQNIITA-SRATQYNHD-----VVKEREYNKLERLHQLVMNKKDKNIAMD--VLNY 231
Db 1238 ---SLQKIRAESEAKQYRRELETIIVREKEAAREL-ERVROLTIEAAKRAAAVEENLLNF 1293
QY 232 VGRADGKRGSWRTDKTEARNEDBMYKILLANDVEYRQKQILME-----NAELEKVKVLOQ 283
Db 1294 RNQLEBNTFRITLEDHLKRD-----LSUNDLE-QQNKMLMBELRRKRNEEELLKLIQ 1348
QY 284 MKKEMI-----SLLSPQ 295
Db 1349 MEKDLAFQKQVABKQLEKQKIELEARRKI TEIQTTCRENALPVCPIQTQATSCRAVTGLQ 1408
QY 296 KKKPRERAED-----GTGTVAISDIEDDSCGELSRDSVWGLSCDVTREQLTNSIRKQWRIL 350
Db 1409 QSHDKQKABELKQOVDELTAANRKAQODMRELTYE-----LNALQLEKTSSEKA-RLL 1461
QY 351 KSHVEKLDN-----QASKVHSEGLNE-----EDVIS-----R 377
Db 1462 KDKLDETNTLRLCLKLELERKQDAEKGYQQRLRQLNQTTGKAEEAMQASDLKKIK 1521
QY 378 ODHEQETEKLELE-----IERCKEMIKAAQOQLLOQQLATTCDDDTTSLLRDCYLL 427
Db 1522 RNYQLELESINHEKGLQREVDRIITRAHAVA EKNIQHLNSQIHSFDEKELERLQICQ-- 1579
QY 428 EKERLKEEWTLFKEQ-KKNFERRRSFTEAAIRLGLERKAFEEERASVWVKQOFLMNTNF 486
Db 1580 RKSDHLKEQFEKSHEQLLQNIKAEN-NDKTQRLNEELEK-SNECAEMLKQKVEELT-- 1635
QY 487 DHONSENVKLFSAFSGSSDPDNLIVHSRPROKKLHSV---ANGVPACTSKLTKSLPASPS 543
Db 1636 -RQNNETKLMQRIQAES--ENIVLEKQTIQORCALKIQADGPKDQLRSTNEHLHKQTK 1692
QY 544 T-SDFRQTHSCVSEHSSISVLNITPEESKPSVARESTDQKWSVQS 588
Db 1693 TEQDFQRIKICLEE-----DLAKSQNLVSEFKQKCDQONIIQN 1731
```

Search completed: June 12, 2006, 19:16:08

Job time : 56 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 13, 2006, 23:29:14 ; Search time 304 Seconds  
(without alignments)  
5677.954 Million cell updates/sec

Title: US-10-644-084-2  
Perfect score: 3165  
Sequence: 1 MGDMMVTDPVLCTENKNLS.....CYSGCSSAFSAHGDRDDLP 615

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1403666 seqs, 93554401 residues  
Total number of hits satisfying chosen parameters: 2807332

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q=/abs/ABSSWEB spool/US10644084/runat 12062006 150424 1461/app query.fasta\_1  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs03p -USER=US10644084 @CGN\_1\_1 307 @runat 12062006 150424 1461  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*  
1: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/1\_COMB.seq.\*  
2: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/5\_COMB.seq.\*  
3: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/6A\_COMB.seq.\*  
4: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/6B\_COMB.seq.\*  
5: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/7\_COMB.seq.\*  
6: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/H\_COMB.seq.\*  
7: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/PP\_COMB.seq.\*  
8: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/RE\_COMB.seq.\*  
9: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/RE\_COMB.seq.\*  
10: /EMC\_Celeerra\_SIDS3/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2720	85.9	2272	3	US-10-104-047-1497
2	635.5	20.1	428	3	US-09-513-999C-3004
3	223.5	7.1	10300	3	US-09-949-016-636
4	222.5	7.0	2718	3	US-09-248-796A-2230
5	222.5	7.0	8590	3	US-09-949-016-5562
6	222.5	7.0	38575	3	US-09-949-016-17304
7	222.5	7.0	119153	3	US-09-949-016-12378
8	220.5	7.0	6276	3	US-09-949-016-2009

9	220.5	7.0	6284	3	US-09-949-016-1028	Sequence 1028, Ap
10	217	6.9	6921	3	US-09-643-597-117	Sequence 117, App
11	217	6.9	6921	3	US-09-480-884A-117	Sequence 117, App
12	217	6.9	6921	3	US-09-542-615A-117	Sequence 117, App
13	217	6.9	6921	3	US-09-606-421B-117	Sequence 117, App
14	217	6.9	6921	3	US-09-221-107-117	Sequence 117, App
15	217	6.9	6921	3	US-09-466-396A-117	Sequence 117, App
16	217	6.9	6921	3	US-09-476-496A-117	Sequence 117, App
17	217	6.9	6921	3	US-09-630-940B-117	Sequence 117, App
18	217	6.9	6921	3	US-09-285-479-117	Sequence 117, App
19	217	6.9	6921	3	US-10-007-700-117	Sequence 117, App
20	217	6.9	7045	3	US-09-919-172-28	Sequence 28, Appl
21	215	6.8	8930	3	US-09-814-915A-91	Sequence 91, Appl
22	212.5	6.7	8257	3	US-09-595-684B-30	Sequence 30, Appl
23	212.5	6.7	8503	3	US-09-620-312D-130	Sequence 130, App
24	211.5	6.7	5053	2	US-08-685-576-2	Sequence 2, Appli
25	210.5	6.7	4065	3	US-09-016-434-1105	Sequence 1105, Ap
26	210.5	6.7	4739	3	US-08-685-871-1	Sequence 1, Appli
27	209.5	6.6	5857	2	US-09-220-132-79	Sequence 79, Appl
28	208.5	6.6	4363	2	US-08-685-576-5	Sequence 5, Appli
29	208.5	6.6	4848	3	US-09-976-594-295	Sequence 295, App
30	208	6.6	2728	3	US-09-799-451-62	Sequence 62, Appl
31	206.5	6.5	5813	3	US-09-949-016-1739	Sequence 1739, Ap
32	206.5	6.5	7500	3	US-09-949-016-4957	Sequence 4957, Ap
33	203.5	6.4	6481	3	US-09-620-312D-1076	Sequence 1076, Ap
34	203	6.4	4747	3	US-09-854-133-729	Sequence 729, App
35	203	6.4	7218	3	US-09-949-016-1775	Sequence 1775, Ap
36	203	6.4	7218	3	US-09-949-016-1776	Sequence 1776, Ap
37	203	6.4	10136	2	US-08-353-700-2	Sequence 2, Appli
38	203	6.4	10136	7	PCT-US95-16216-2	Sequence 2, Appli
39	203	6.4	82125	3	US-09-949-016-13517	Sequence 13517, A
40	203	6.4	82125	3	US-09-949-016-13518	Sequence 13518, A
41	202	6.4	5877	3	US-10-028-946-3	Sequence 3, Appli
42	202	6.4	6165	3	US-10-028-946-1	Sequence 1, Appli
43	202	6.4	6189	3	US-09-964-958-10	Sequence 10, Appl
44	202	6.4	6201	3	US-09-964-958-8	Sequence 8, Appli
45	201	6.4	6159	3	US-10-017-216-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-10-104-047-1497  
; Sequence 1497, Application US/10104047  
; Patent No. 6943241  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. 6943241el full length cdNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1497  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1497

Alignment Scores:  
Pred. No.: 2.48e-284 Length: 2272  
Score: 2720.00 Matches: 536  
Percent Similarity: 86.8% Conservativeness: 34  
Best Local Similarity: 81.6% Mismatches: 40  
Query Match: 85.9% Indels: 47  
DB: 3 Gaps: 3

US-10-644-084-2 (1-615) x US-10-104-047-1497 (1-2272)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
|||||  
|||||

Db 224 ATGGGAGATTGGATGACTGTTACAGATCCA-----GAAAGCAAAACTATCTCT 271  
Qy 21 GlnTyrThrSerGlnThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
Db 272 CAATATACCTCAGAAACAAGATGCTCCATCAAGTTTATATCTCAGCAGCAAGTGCTATGT 331  
Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
Db 332 TCTTCAATACCTTTATCGAAATATGTCAGAGTTTTCAGTGCCTCTCTGCAGAGAT 391  
Qy 61 AsnIleGlnSerIleSerTyrLeuAspGlnGlnLeuThrThrPheGlyPheProSer 80  
Db 392 AATATTTGAACAGAGTATCTCATATCTTGATCAGGAATGACTACTTTTGGTTTCTTCA 451  
Qy 81 LeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
Db 452 TTATATGAAGAATCCAAAGGTAAAGAGACAAGAGAGAGATTAAATATATAGTCTGACTPA 511  
Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
Db 512 AATTGTATGAATGAGCTGTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAATGTG 571  
Qy 121 GlnThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
Db 572 GAGACACAGAAATTTGAAGCGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAA 631  
Qy 141 LeuLysGlnGlnLeuGlnThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
Db 632 CTTAAGGAACAACCTGGAAACCTCCAGAGGGAATGATTGGGCTTCAGGAAAGAGACAGA 691  
Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluVal 180  
Db 692 CAGTTACATGTAGAACAGGAATTTGCATCAGCTACTAAAGATGAGAAAGATGAGGTG 751  
Qy 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
Db 752 CAAAATTTACAAATATCATTCGACGTCGAGCTACTCAGTAAATCATGATGAAGAGA 811  
Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
Db 812 AAAGAGCGGTGAATATAATAACTGAAGGAACGTCTACATCAACTGTTTATGAACAGAAA 871  
Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db 872 GATAAGAAATAGCTATGACATTTGTAATTTATGTCGGCAGAGCTGATGNAAGAGGC 931  
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Db 932 TCCTGGAGGACTGGTAAAACTGAAGCCAGGAATGAAGATGAATGTATAAAATCTCTTTG 991  
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280  
Db 992 AATGATTATGAATATCGTCAGAAACAAATTCCTAATGGAATATGCAAGACTTAAGAAGTT 1051  
Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
Db 1052 CTTCAACAATGAAAGGAATGATTTCTTCTTCTCCCAAAAGAAAGAACTTAGA 1111  
Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGly 320  
Db 1112 GAAAGAGTAGTAGTAGTACAGGAACCTGTT---ATTTCGATGTTTGAAGAAGATCCGGG 1168  
Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGlnGlnLeuThr 340  
Db 1169 GAACTAAGCAGAGAGAGTAGTAGTGGACCTTTCTGTGAACCTGTGAGAGAGCGCTTACA 1228  
Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
Db 1229 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAGCTTGAATACCA 1288  
Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
Db 1289 GTTTCAAAGGTACACCTGGAGGTTTAAATGATGAAGATGTAAATCTCAGCAAGACCAT 1348

Qy 381 GluGlnGlnThrGluLysLeuGluLeuGluIleGluArgCysValGluMetIleLysAla 400  
Db 1349 GAACAAGAAACTGAAAAACTCGAGTTAGAAATTCAGCAGATGTAAAGAAATGATTAAGACT 1408  
Qy 401 GlnGlnGlnLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
Db 1409 CAGCAACAGCTTTTACACAGCAGCTCGCTACTGCTATGATGATGATACCACTTCACTA 1468  
Qy 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLysGluLysGluLysGluLysTrpThrLeuPhe 440  
Db 1469 TTACGAGACTGTTATTTGTTGGNAAGAAAGCAAGCTCTCAAAGAAAGAAATGTCCTCTTTT 1528  
Qy 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460  
Db 1529 AAAGACAGAAAAGAAATTTTTCAGAGGAGAGACGAGCTTTTACAGAGCGCTATTTCG 1588  
Qy 461 LeuGlyLeuGlu----- 464  
Db 1589 CTGGGATTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACACAGAAATCTCTTTTGGGT 1648  
Qy 464 ----- 464  
Db 1649 ATCCATCATTCATCCACGGTCAGTAGTGTGCTGTCAGAAAGACAGTGCACCAAGACTAT 1708  
Qy 465 -----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnG1 479  
Db 1709 ATTTGCTGTTTAAAAAGAAAGGCATTTTGAAGAGAAAGAGCCAGTTCGTTAAAGCAGCA 1768  
Qy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499  
Db 1769 GTTTCCTAAATATGACTACCTTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTTCAGTGC 1828  
Qy 499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLys 519  
Db 1829 CTTCTCAGGAAGTCTGATTGGGACAACTTATATAGTGCACTCGAGCGCAGCCGCAAAAGAA 1888  
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539  
Db 1889 GCCTCAGAGTGTGCTAATGGGCTCTCCAGTTTGCATGTCTAAACTTACTAACTCTCTCC 1948  
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559  
Db 1949 TGCTTCACCTCCACTTCAGACTTTTGCACACAGCTTCCTGCATATCTGAACATAGTTC 2008  
Qy 559 rLisSerValLeuAsnIleThrProGluLysSerLysProSerGluValAlaArgGluSe 579  
Db 2009 AATCAATGTACTGAATATACTGCTGAAGAAATTAACCAAAATCAGGTTCGAGGAGAACG 2068  
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerG1 599  
Db 2069 TACAAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCAGAGGAAGGTTGCTATAGTGG 2128  
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspAspAspLeuPro 615  
Db 2129 ATGCTCCTTGAGCTACACAAATTCATGTAGAAAAAGATGACTTACCT 2177

## RESULT 2

US-09-513-999C-3004  
; Sequence 3004, Application US/09513999C  
; Patent No. 6783961  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Duclert, A.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
; Patent No. 6783961  
; FILE REFERENCE: 59 US2. REG  
; CURRENT APPLICATION NUMBER: US/09/513,999C  
; CURRENT FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/122,487  
; PRIOR FILING DATE: 1999-02-26  
; NUMBER OF SEQ ID NOS: 36681



QY 279 -----LysValLeuGlnMetLysValGluMetIleSerLeuLeuSerProGlnLys 296  
DB 4546 GAAAGTAGAACAAAGCAACAAATACAAAGAACTGCAAGCTGCCCTTATTTCCCGAATA 4605  
QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrVal 310  
DB 4606 GAAGCACTAAAGAAACAAAGTCTCCAAGAGGAATTCCTTTGGCCAGAGTACCAT 4665  
QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323  
DB 4666 GAACGCTCTCAACAGTCTCTGCAGATGTGGAAGCCAAAGTTTCTGCTCAAAATAAGAA 4725  
QY 324 ArgAspSerValTrpGly----- 329  
DB 4726 AAAGATACGGTCTTAGGAAGGTTAGCTCTTCTTCAAGAAAGAGACAAACTCAATTACA 4785  
QY 330 -----LeuSerCysAspThrValArg 336  
DB 4786 GAAATGGACAGTCTTTATTGGAAATCAGAGTCTCAGCAGCTCCTGTGAAAGTCTAAA 4845  
QY 337 -----GluGlnLeu 339  
DB 4846 CTAGCTCTAGAGGCTTCTACTGAAGCAAGAAAGTTAGTGAAGAAATTTGAATCTTTG 4905  
QY 340 ThrAsnSer-----IleArgLys 345  
DB 4906 AAATCTCTAAGATTGCACAAAGTACTGAGTGGCAAGAAACACACAGGAGTACAAAA 4965  
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365  
DB 4966 GAGTATGAAATCTCTCGAGTCTTATGAGATGTTAGTAAATGAAGCAGAAAGGATTGAG 5025  
QY 366 Ser-----GluGlyLeuAsnGlu-----AspValIleSerArg 377  
DB 5026 CATGTGTCGAAGCTGTGAGGCAAGAGAAACAAAGAACTGTATGGCAAGTTAAGAGCACA 5085  
QY 378 GlnAspHisGluGlnGluThrGluLys-----LeuGluLeuIleGluArg 393  
DB 5086 GAGGCAACAGAAAGAGAGACAGAAAGACAGTTCGAGGAAGCTGACAGAAATGAGGNA 5145  
QY 394 CysLysGluMetIleLys-----AlaGlnGlnGlnLeuGluGlnGln 408  
DB 5146 ATGAAAGAAAGATGAGAAAGTTTGCTAAATCTAAACAGCAGAAATCTCAGAGCTGGA 5205  
QY 409 -----LeuAlaThrThrCys 413  
DB 5206 GAAGAGAATGACCGCTTAGGCAGAGGTGCACCTGCAGAGATACAGCTAAAGAGTGT 5265  
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysThrLeuLeuGluGluLysGluArgLeu 433  
DB 5266 ATGGAA-----ACACTCTTTCTTCCATGCGCAGCATGAAGAGAACTTGAAGGGTC 5319  
QY 434 LysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgSer 453  
DB 5320 AAAATGAGTATGAAACCTTTCTAAGAAAGTTTCAGTCTTTAATGCTGAGAAAGACTCT 5379  
QY 454 PheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluArgAla 473  
DB 5380 CTAAGTGAAGAGGTTCAAGATTTA-----AAGCATCAGATGAAGATAATGTA 5427  
QY 474 SerTrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsn 493  
DB 5428 TCT-----AAACAGCTAACCTAGAGGCCACCGAG-----AAACATCATNACCAACGAAAT 5478  
QY 494 ValLys-----LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIle 510  
DB 5479 GTCACTGAAGAGGAAACAGTCTATACCAGGTGAGACTGAAGAGCAAGACTCTCTGAGT 5538  
QY 511 ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys 530  
DB 5539 ATGAGCACAAAGA-----CCTACATGT 5559

QY 531 ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSer---AspPheArgGln 549  
DB 5560 TCAGNATCGGTTCATCATCAGCCAGAGTCCCAACCTCGTGTAAAGTAAGGATTTC----- 5613  
QY 550 ThrHisSerCysValSerGluHisSerSerIleSer----- 561  
DB 5614 -----AGCTCATCATGATAAATTAATAACTTACTACAGCAGATTGATCAG 5658  
QY 562 -----ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArg 577  
DB 5659 CTCAAAGAAAGAAATTCGTGGATTAGAGGAGGAGAGCAAAAAACAAGGAATTTAGCCAG 5718  
QY 578 GluSerThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGly 595  
DB 5719 ACTTTAGAAATGAGAAAAATACCTTACTGAGTCAGATATCAACAAAGGATGTT 5772  
RESULT 4  
US-09-248-796A-2230  
; Sequence 2230, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIORITY FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 2230  
; LENGTH: 2718  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-2230  
Alignment Scores:  
Pred. No.: 2,53e-13 Length: 2718  
Score: 222.50 Matches: 137  
Percent Similarity: 36.5% Conservative: 140  
Best Local Similarity: 18.1% Mismatches: 264  
Query Match: 7.0% Indels: 217  
DB: 3 Gaps: 27  
US-10-644-084-2 (1-615) x US-09-248-796A-2230 (1-2718)  
QY 10 ProValLeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSer 29  
DB 427 CCTATTTTA-----CTAGATGCATATAAAATCGAGAAGAAGAAGAT 468  
QY 30 ProSerSerLeuTyrSerGlnValLeuCysSerSerValProLeuSerLysAsnVal 49  
DB 469 GAAATAATTAGCATATTCGAAAAAGTATTACAAGAAATATACACAGTCAGTTCATTCGTCA 528  
QY 50 HisGlyValPheGlyVal---PheCysThrGlyGluAsn-----IleGlu 63  
DB 529 GAAGTAAACTATTAGTTTGAATTTGAATGGTCTTAATGAGAACACAAAATTACTCAAT 588  
QY 64 GlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyrGlu 83  
DB 599 GGTGAGATTAGAAAGTTGACTTAAAGATAATAATACCATGAACCTTCAAAATAGTT----- 642  
QY 84 GluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMet 103  
DB 643 -----AAATTAGAGGAAGTAATAAAGAACTTAAC-----AACGACGTC 681  
QY 104 AsnGluLeuLeuValLeuGlnArg---LysAsnLeuLeuAlaGln----- 117  
DB 682 AAACGATTAGATACGATCCAAAGGAAAGAAAGATCATGGGGAAGAAAGATTGATTCAATG 741  
QY 118 -----GluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHis 133



```
Db 742 AGTGAAGAAATTGGAAATAATCAATAATCAGAAATTCAGAAATTCAGAAAGACCTTTAAAGAC 801
QY 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuThrSerArgArgGluMetIle 153
Db 802 AAAGAG-----GAAAAAATAGAGAGCTTGAAGCTTGAGATAACT 840
QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
Db 841 GATCGTGATACAGAAATTAAGATCTTTGAGTCTTAATACATCTAGTAAACAGCAAGGAATTA 900
QY 174 LysAsnGlnLysAspGluValGlnLysLeu----- 183
Db 901 AATACAGAATTGGAAGAGCCCAAGAGGTGCCAACACAGAAAAGGTCATCAAGATAAGTCT 960
QY 184 -----Gln 184
Db 961 TCTTCTGGAGAAGAAACACTGATAAAGATAGTAGCGTGGTGATAGCAGCGACAATCT 1020
QY 185 AsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGlu 204
Db 1021 GATGTGTGTAATCTGAAATTTCTCAATTGAAATCCCAATTATCTACCAAGAAACCGAG 1080
QY 205 TyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsn--- 223
Db 1081 GTTGAAGAATTGACTAACGAAGTCAGNACCATTTAAAGAGTCAATTGAACGATAAAGATGAG 1140
QY 224 -----IleAlaMetAspValLeuAsnTyrValGlyArgAlaAsp 236
Db 1141 GAGATAGAAGATTTACGAGACCTGGTTAAGGAAATAGGCAATGAGCTAGTAACTCTTAAG 1200
QY 237 GlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyr 256
Db 1201 GATGAAATCAAGTCTCTAAAGAAATCTCAAAATATCCACTGACATGAAGATAGTACCACC 1260
QY 257 Lys-----IleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
Db 1261 AAGAAGATACTAATACTCAATAATATGATTGGGAATTTGAATCAATCTCAGTTAAAA 1320
QY 273 GluAsnAlaGluLeuLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeu 292
Db 1321 GAAAAATAAAGAAATTGAAGAAGAATTGGACGTGGCTACAAAAGAA----- 1365
QY 293 SerProGlnLysLysLysProArgGluArgAlaGluAsp----- 305
Db 1366 TCTAAAGAAAAAGAAAAAGAACGCAAGAAATCAGNAGATCAATTGAGAAAACAGATCCAA 1425
QY 306 -----GlyThrGlyThrValAlaIle 312
Db 1426 ACTTTGAAAACAAAACTCGAATCAACAGAAAAAATTTGAATCCAGACCAAGGATTTA 1485
QY 313 SerAspIleGluAspAspSerGlyGluLeu----- 322
Db 1486 AAGAACCTTGCTGGAAGAGAAAGAAAAAATTAGAGAAGCGAATCTCAGAAATTTGTCAAAGTTC 1545
QY 323 ---SerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsn 341
Db 1546 AAAAGTAATGATTTCATCTTTGAAATTAGAAATTTCTTCATTTGAAATTCCTCATTAACCAAT 1605
QY 342 -----SerIleArgLysGlnTrpArg 348
Db 1606 AAGATAAATTCATCAACGAACTAAAGACTCAAGTCGATGATGTATCAACAGCTCAATAAAA 1665
QY 349 IleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGly 368
Db 1666 TCATTGAACCTCTAAAGTTGAAGAGTTGTCAAAATCTTAATAACGAGTTCGAGTCCCAATTCT 1725
QY 369 Leu-----AsnGluGluAspValIleSerArgGlnAsp----- 379
Db 1726 ATGGATTTTTTGAAGACAAAAATGAATTTGTTAAACCAAAACAAAGAGTATTAAATGGAATAAT 1785
QY 380 -----HisGluGlnGluThrGluLysLeuGluLeu 389
```

```
Db 1786 ACAAGATCATTAACACCTGCAGTTACAAAATTCGAAACAGAGAAAGATGTAATTACT 1845
QY 390 GluIleGluArgCysLysGlu----- 396
Db 1846 GAGTTGGAAAAAGACGAAACAACTAGACATTTGTGATTTCAGATAAATCACAATCGCA 1905
QY 397 -----MetIleLysAlaGlnGlnGln 403
Db 1906 AACGATATGTTATCTTTCAAAAAAGCAACATGAAGAATTTGATGATGAAGTCTCAAGAGTAC 1965
QY 404 LeuLeuGlnGlnGlnLeuAlaThrCysAspAspThrThrSerLeuLeuArgAsp 423
Db 1966 AGTTTACGAAATGAA-----AGTTTGGAGATGATTTTGACTGAAGCCAGAAAT--- 2013
QY 424 CysTyrLeuLeuGluGluLysGluArg-----LeuLysGlu 435
Db 2014 -----TTGTTTCAACAGAGAGAACCCGCGAGACTTCAAATATATGAGAAGTTATTAGTTGAC 2067
QY 436 GluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThr 455
Db 2068 GCGAGAGAAATGTTGAAACAACAAAGCAGGATTCTTAAGCTTGAA----- 2112
QY 456 GluAlaAlaIleArgLeuGlyLeuGluArgLysAla---PheGluGluGluArgAlaSer 474
Db 2113 -----ATTCTAGACATTTGGAGGATTAAGCAGAGATTCGAAAGAAATAATTCGCTCT 2163
QY 475 TrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHis----- 488
Db 2164 TTAATAAAGAGGAACACAGAGAACTAGATGAATTAAGAAATCAGTTCAAGAGTACAAA 2223
QY 489 -----GlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAsp 505
Db 2224 TTGAAAGTCGAAAAATTTGGAANAAC---AAAGTTATGACTCTCGAAAAAGAAAAAGAGC 2280
QY 506 ProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsn 525
Db 2281 AAGGAAAGTTTAAAGCACCATTAACCAACCAATCAGTAAA-----GAGCTATCATCG 2334
QY 526 GlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSer 545
Db 2335 ACCATTGAAACGTTGAGAACCGCATTTGAATAATTCACCAATAAATAATAGAGATCTTGAA 2394
QY 546 AspPheArgGlnThrHisSerCysValSerGluHisSerSerIle----- 560
Db 2395 AATTATATAATAAACCTTAAGAAATTTAGTGAAGACAACACATTAAGATTTGAACGTTTA 2454
QY 561 -----SerValLeuAsnIleThrProGluGluSerLysProSerGluValAla 576
Db 2455 TCCAAAAAATAATAAATTTTGAATCAACAGTATCAGTATCAGTAAAGACGTTAAAGTTCT 2514
QY 577 ArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGlu 594
Db 2515 GAATCATCAACAAAGCAGGATAAACCCAGTTGAGACTGCACCTCGAAAGCAACGAT 2568
```

## RESULT 5

```
US-09-949-016-5562
; Sequence 5562, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

; SEQ ID NO 5562  
 ; LENGTH: 8590  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-09-949-016-5562

## Alignment Scores:

Pred. No.: 1,77e-12 Length: 8590  
 Score: 222.50 Matches: 139  
 Percent Similarity: 38.9% Conservative: 147  
 Best Local Similarity: 18.9% Mismatches: 217  
 Query Match: 7.0% Indels: 233  
 DB: 3 Gaps: 32

US-10-644-084-2 (1-615) x US-09-949-016-5562 (1-8590)

QY 15 GluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSer----- 29  
 DB 2097 GAAATGAGATATTGGAGACCAGCTAAGGCAACTCCAGATTCAGTAAGGAATCCATA 2156  
 QY 30 -----ProSerSerLeuTyrSerGlnGlnValLeuCysSerSerValPro--- 44  
 DB 2157 GACGGAAAACTCCCAAGCACA-----GACCAGCAGGAATCGTGTCTTCCACTCCAGGT 2210  
 QY 45 LeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGln 64  
 DB 2211 TTAGAAGAACCTTTATTCAAGCCACAGACAGCATCACATCAACCTGTTTAGAGTCC 2270  
 QY 65 SerIle-----SerTyrLeuAspGlnGlnLeuThrThrPheGlyPhePro 79  
 DB 2271 AACTTGTGCCAGACTGGCGCTTCTCATCTGAAGATGCGAGTGTCTGCGAGCGCGAACT 2330  
 QY 80 SerLeuTyrGlu-----GluSerLysSerLysGluAlaIysArgGluLeuAsnIle 96  
 DB 2331 TCTGTGCCAGATTAAAGCCAGCTGAAGGAAATAGAGCGCTGAGAAAGTAGAGTTAGAA 2390  
 QY 97 ValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAla 116  
 DB 2391 TTGAAGATTAGTCTACACAACTGAGCTTACTAAAAATCAGAGAGGTATTTCAGTTA 2450  
 QY 117 GlnGluSerValGluThrGlnAsnLeuLysLeuGlySer-----AspMetAspHisLeu 134  
 DB 2451 CAAGAGCAGATAAATAACAGAGGTTTAGAAATCGAGAGTCTAAAGACAGATATCCCATGAA 2510  
 QY 135 GlnSerCysTyrAlaLys--LeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153  
 DB 2511 GCTGAAGTCCATGCCCAAGCTGCGACGAGAAATGGAAAGCAGCCCACTACAAATTGCT 2570  
 QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeu 173  
 DB 2571 GGCCTA--GAACATCTAAGAGATTGCAACCTAAA----- 2603  
 QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGln 193  
 DB 2604 -----CTGGATCAACTGCAAAAACCTC----- 2624  
 QY 194 TyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHis 213  
 DB 2625 -----ATAAGCAAAAGGAAGAGAGCTTTAGTACTCTTCCTGGACAACT--- 2669  
 QY 214 GlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGly 233  
 DB 2670 -----AGTGAGAAAGAGCAGCTCTCACTAAAAATA----- 2699  
 QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253  
 DB 2700 -----CAGACAGAGATAATAGAACCAAGAA 2723  
 QY 254 GluMetTyrLysIleLeu-----LeuAsnAspTyrGluTyrArg 266  
 DB 2724 GATTTAATTAGCTCTGCATACAGCTAGAAAATGCAAGCCCAAGAGCATGATGAGAGG 2783  
 QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys----- 278

DB 2784 ATAAAGCAGCTACAGGTGGAACCTTTGTGAAATGAAGCAAAACCAGAAAGAGATTGGAAA 2843  
 QY 279 -----LysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296  
 DB 2844 GAAAGTAGACAAAGCAACAATACAAAGGAAACTGCAAGCTGCCCTTTATTTCCCGAAAA 2903  
 QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310  
 DB 2904 GAAGCACTAAAGAAAAACAAGTCTCCAAGAGGAATGTCTTTGGCCAGAGTACCAT 2963  
 QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323  
 DB 2964 GAACGTCTCACCAAGTCTCTGCGACATGTGGAAGCCCAAGTTTCTGCTCAAAATAAGAA 3023  
 QY 324 ArgAspSerValTrpGly----- 329  
 DB 3024 AAGATACGGTCTTAGGAAAGGTAGTCTCTTCAAGAAGAAAGAGACAAACTCATTA 3083  
 QY 330 -----LeuSerCysAspThrValArg 336  
 DB 3084 GAAATGGACAGGTCTTTATTGGAAAAATCAGAGTCTCAGCAGCTCTGTGAAAGTCTAAA 3143  
 QY 337 -----GluGlnLeu 339  
 DB 3144 CTAGCTCTAGAGGTCTTACTGAAGCAAGGAAAGTTAGTGAAGGAAATCGAATCTTTG 3203  
 QY 340 ThrAsnSer-----IleArgLys 345  
 DB 3204 AAATCTTTAAGATTGCAAGAAAGTACTAGTGGCAAGAGAAACAACAAGAGAGCTACAAA 3263  
 QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365  
 DB 3264 GAGTATGAAATCTTCTCAGCTCTATGAGAAATGTTAGTAATGAAGCAGAAAGGATTCAG 3323  
 QY 366 Ser-----GluGlyLeuAsnGluGlu-----AspValIleSerArg 377  
 DB 3324 CATGTGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGGCAAGTTAAGAAAGCACA 3383  
 QY 378 GlnAspHisGluGlnGluThrGluLys-----LeuGluLeuGluLysGluArg 393  
 DB 3384 GAGGCAAAACAAGAGGAGACAGAAAGAGCAGTTGCAAGAGCTGAGCAAGAAATGGAGGAA 3443  
 QY 394 CysLysGluMetIleLys-----AlaGlnGlnGlnLeuGlnGln 408  
 DB 3444 ATGAAAGAAAGATGAGAAAGTTTGCTAAATCTAAACAGCAGAGAAATCTTAGAGCTGGA 3503  
 QY 409 -----LeuAlaThrThrCys 413  
 DB 3504 GAAGAGAATGACCGGCTTTAGGGCAGAGGTGCACCTCGCAGGAGATACAGCTAAAGAGTGT 3563  
 QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluLysGluArgLeu 433  
 DB 3564 ATGGAA-----ACACTTCTTTTCCCAATGCCCATGAAGGAAAGAACTTGAAGAGGTC 3617  
 QY 434 LysGluGluTrpThrLeuPheLysGluLysLysAsnPheGluArgArgSer 453  
 DB 3618 AAAATGGAGTATGAACCCCTTTCTAAGAAGTTTTCAGTCTTTTAATGCTCTGAGAAAGCTCT 3677  
 QY 454 PheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAla 473  
 DB 3678 CTAAGTGAAGAGGTTCAAGATTTAAAGCATCAGATAGAGGTTAATGTATCTTAAACAAGCT 3737  
 QY 474 SerTrp---ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492  
 DB 3738 AACCTAGAGGCCACCGAAGAACATGATAACCAACCAATGTCTCCTGAAGAGGGAACACAG 3797  
 QY 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512  
 DB 3798 TCTATACCA-----GGTCAGAGCTGAAGAGCAAGACTCTCTGATGATGAGC 3842  
 QY 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532

Db 3843 ACAAGA-----CCTACATGTTTCAGAA 3863  
 QY 533 LysLeuThrLysSerLeuProAlaSerProSerThrSer---AspPheArgGlnThrHis 551  
 Db 3864 TCGGTTCCATCAACGAGAGTGCACACCTGCTGTAAGTAAGGATTTTC-----3911  
 QY 552 SerCysValSerGluHisSerSerIleSer-----561  
 Db 3912 -----AGTCACATGATGAATTAATACTACTACAGCAGATTTGATCAGCTCAAA 3962  
 QY 562 -----ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSer 579  
 Db 3963 GAAAGAAATTCGTGATTAGAGGAGGAGACGACAGAAAAACAAGGAATTTAGCCAGACTTTA 4022  
 QY 580 ThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGly 595  
 Db 4023 GAAATGAGAAAAATACCTTACTGAGTCAGATATCAACAAAGGATGGT 4070  
 RESULT 6  
 US-09-949-016-17304  
 ; Sequence 17304, Application US/09949016  
 ; Patent No. 6812339  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
 ; FILE REFERENCE: CLO01307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17304  
 ; LENGTH: 38575  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)...(38575)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-949-016-17304  
 Alignment Scores:  
 Pred. No.: 2,25e-11 Length: 38575  
 Score: 222.50 Matches: 139  
 Percent Similarity: 38.9% Conservative: 147  
 Best Local Similarity: 18.9% Mismatches: 217  
 Query Match: 7.0% Indels: 233  
 DB: 3 Gaps: 32  
 US-10-644-084-2 (1-615) x US-09-949-016-17304 (1-38575)  
 QY 15 GluAsnLysAsnLeuSerGlnTyThrSerGluThrLysMetSer-----29  
 Db 4097 GAAATGGAATATTGGACACAGCTTAAGGCAACTCCAGATTCAGTAAGGGAATCCATA 4156  
 QY 30 -----ProSerSerLeuTyThrSerGlnGlnValLeuCysSerSerValPro---44  
 Db 4157 GACGGAAAACTCCCAAGCACA-----GACCAGCAGGAATCGTGTCTTCCACTCCAGGT 4210  
 QY 45 LeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGln 64  
 Db 4211 TTAGAGAACCTTTATTCAAGCCACAGACAGATCACACTCAACCTGTTTATAGAGTCC 4270  
 QY 65 SerIle-----SerTyLeuAspGlnGluLeuThrThrPheGlyPhePro 79  
 Db 4271 AACTTGTGCCAGACGTGGCCTTCTCATTTCTGAAGATGCGAGTGCTCTGCGAGGCGGAACT 4330

QY 80 SerLeuTyTrpGlu-----GluSerLysSerLysGluAlaLysArgGluLeuAsnIle 96  
 Db 4331 TCTGTTGCCAGCATTAAGCCAGCTGAAGGAATAATAGAGGCTGAGAAAGTAGAGTTAGAA 4390  
 QY 97 ValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAla 116  
 Db 4391 TTGAAAGTTAGTTCTTACAACAAGTGAGCTTACTAAAAAATCAGAAAGGATTTTCAGTTA 4450  
 QY 117 GlnGluSerValGluThrGlnAsnLeuLysLeuGlySer-----AspMetAspHisLeu 134  
 Db 4451 CAAGACAGATAAATAACAGAGGTTTAGAAAATCGAGAGTCTAAAGACAGATATCCCATGAA 4510  
 QY 135 GlnSerCysTyAlaLys---LeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153  
 Db 4511 GCTGAAGTCCATGCCGAAAGCCTGCAGCAAAATTTGGAAGCAGCAACATACAAATTGCT 4570  
 QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeu 173  
 Db 4571 GGCCTA--GAACATCTAAGAGAAATTGCAACCTTAA-----4603  
 QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGln 193  
 Db 4604 -----CTGGATGAACCTGCAAAAACCTC-----4624  
 QY 194 TyrAsnHisAspValLysArgLysGluArgGluTyThrAsnLysLeuLysGluArgLeuHis 213  
 Db 4625 -----ATAAGCAAAAAGGAAGACGTTAGCTACTCTTTCTCGACAACCTT---4669  
 QY 214 GlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyValGly 233  
 Db 4670 -----AGTGAGAAAGAGCAGCTCTCCTCAATAATA-----4699  
 QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253  
 Db 4700 -----CAGACAGAGATAATAAGAACAGAA-----4723  
 QY 254 GluMetTyTrpLysIleLeu-----LeuAsnAspTyTrpGluTyArg 266  
 Db 4724 GATTTAATTAAGGCTCTGCATACAGCTAGAAAATGCAAGCCAAAGAGCATGATGAGAGG 4783  
 QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys-----278  
 Db 4784 ATAAAGCAGCTACAGGTGGAACTTTGTGNAATGAAGCAAAAACAGAGAGATGAGAGAA 4843  
 QY 279 -----LysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296  
 Db 4844 GAAAGTAGACAAAGCAACAATAACAAGGAAACTGCAAGCTGCCCTTATTTCCCGAAAA 4903  
 QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310  
 Db 4904 GAAGCACTAAAGAAAAACAAAAGTCTCCAAGAGGAATTTGTCTTTGGCCAGAGGTACCAT 4963  
 QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323  
 Db 4964 GAACGTCTCACCAAGTCTCTGCAGATGTGGAAAGCAAGTCTTCTGCTCAAAATAAGAA 5023  
 QY 324 ArgAspSerValTrpGly-----329  
 Db 5024 AAAGATACGGTCTTAGGAAGGTAGCTCTCTTCAAGAAGAAAGACACAACTCATTACA 5083  
 QY 330 -----LeuSerCysAspThrValArg 336  
 Db 5084 GAAATGGACAGGTCTTTATTGGAAAAATCAGAGTCTCAGCAGCTCTCTGTGAAAGTCTTAAA 5143  
 QY 337 -----GluGlnLeu 339  
 Db 5144 CTAGCTCTAGAGGCTTCTTACTGAACAGGAAGAAAGTAGTGAAGAAATGGAATCTTTG 5203  
 QY 340 ThrAsnSer-----IleArgLys 345  
 Db 5204 AAATCTTCTTAAGATTGCAGAAAGTACTGAGTGGCAAGAGAAACACAGGAGCTACAAAA 5263  
 QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365



```
QY 234 ArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAsp 253
Db 37278 -----CAGACAGAGATAATAGAACCAAGAA 37301
QY 254 GluMetTyrLysIleLeu-----LeuAsnAspTyrGluTyrArg 266
Db 37302 GATTTAATTAAGGCTTCGCATACACAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGG 37361
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLys----- 278
Db 37362 ATAAAGCAGCTACAGGTGAAACTTTGTGAATGAAGCAAAAAACCAGAGAGATTGGAGAA 37421
QY 279 -----LysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLys 296
Db 37422 GAAAGTAGAACCAAGCAACAATACAAAGAACTGCAAGCTGCCCTTATTTCCCGAAAA 37481
QY 297 -----LysLysProArgGluArgAlaGluAspGlyThrGlyThrVal 310
Db 37482 GAAGCACTAAAGAAACAAAGTCTCCAAGAGGAATTGCTTTGGCCAGAGTACCATT 37541
QY 311 -----AlaIleSerAspIleGluAsp-----AspSerGlyGluLeuSer 323
Db 37542 GAACGTCTCACCAGTCTCTGGCAGATGTGGAAGCAAGTTTCTGCTCAAAATAAGAA 37601
QY 324 ArgAspSerValTrpGly----- 329
Db 37602 AAAGATACGGTCTTAGGAAGGTTAGCTCTTCTTCAAGAAAGAGACAAATCTCATTA 37661
QY 330 -----LeuSerCysAspThrValArg 336
Db 37662 GAAATGGCAGGCTTTATTGGAAATACAGTCTCAGCAGCTCTGTGAAAGTCTAAAA 37721
QY 337 -----GluGlnLeu 339
Db 37722 CTAGCTCTAGAGGCTTACTGAAGACAAGAAAGTTAGTGAAGAAATTCGAATCTTTG 37781
QY 340 ThrAsnSer-----IleArgLys 345
Db 37782 AAATCTTTAAGATTGCAGAAAGTACTAGTGGCAGAGAAACACAGAGAGCTACAAAA 37841
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHis 365
Db 37842 GAGTATGAATTTCTTCGAGTCCCTATGAGATGTTAGTAATGAGCAGCAAGAGATTGAG 37901
QY 366 Ser-----GluGlyLeuAsnGluGlu-----AspValIleSerArg 377
Db 37902 CATGTGGTGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGGCAAGTTAAGAGACCA 37961
QY 378 GlnAspHisGluGlnGluThrGluLys-----LeuGluLeuGluIleGluArg 393
Db 37962 GAGGCAACCAAGAGAGAGACAGAAAGAGCAGTTGCGAGAGAGCTGAGCAAGAAATGGAGAA 38021
QY 394 CysLysGluMetIleLys-----AlaGlnGlnGlnLeuLeuGlnGln 408
Db 38022 ATGAAGAAGAAAGATGAGAAGTTTGTCTAAATCTAAACAGCAGCAAAATCTTAGAGCTGAA 38081
QY 409 -----LeuAlaThrThrCys 413
Db 38082 GAAGAGAATGACCGGCTTAGGCGCAGAGGTGCACCTGCAGGAGATACAGCTAAAGAGTGT 38141
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArgLeu 433
Db 38142 ATGGAA-----ACACTCTTTCTTCCCAATGCCAGCATGAAGAGAACTTGAAGAGGTG 38195
QY 434 LysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgSer 453
Db 38196 AAAATGGAGATGAACACCTTTCTAAGAGATTTCAGTCTTTAATGCTGTGAGAAAGACTCT 38255
QY 454 PheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAla 473
Db 38256 CTAAGTGAAGAGGTTCAAGATTTAAAGCATCAGATAGAAGGTGAATGATATCTATCAACAGCT 38315
QY 474 SerTrp-----ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
```

---

```
Db 38316 AACCTAGAGCCACCAGAGAAACATGATAACCAACGAATGTCACTGAAGAGGGAACACAG 38375
QY 493 AsnValLysLeuPheSerAlaPheSerGlySerAspProAspAsnLeuIleValHis 512
Db 38376 TCTATACCA-----GGTGAGACTGAAGACAGACTCTCTGAGTATGAGC 38420
QY 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
Db 38421 ACAAGA-----CCTACATGTTCCAGAA 38441
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThrSer-----AspPheArgGlnThrHis 551
Db 38442 TCGGTTCCATCAGCCAGAGAGTGCACACCTGCTGTAAGTAAGGATTTC----- 38489
QY 552 SerCysValSerGluHisSerIleSer----- 561
Db 38490 -----AGTCACATCATGAAATTAATACTACACAGAGATTGATCAGCTCAA 38540
QY 562 -----ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSer 579
Db 38541 GAAAGAAATTCCTGGATTAGAGGAGGAGAGCAAGAAAAACAAGAAATTTAGCCAGACTTTA 38600
QY 580 ThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGly 595
Db 38601 GAAATGAGAAAAATACCTTACTGAGTCAGATATCAACAAAGGATGTT 38648

RESULT 8
US-09-949-016-2009
; Sequence 2009, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2009
; LENGTH: 6276
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2009

Alignment Scores:
Pred. No.: 1,72e-12 Length: 6276
Score: 220.50 Matches: 157
Percent Similarity: 36.2% Conservative: 118
Best Local Similarity: 20.7% Mismatches: 233
Query Match: 7.0% Indels: 251
DB: 3 Gaps: 36

US-10-644-084-2 (1-615) x US-09-949-016-2009 (1-6276)
QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
Db 3033 AGTAAGCAAGTTAAGAATATCGAATTCGAATTCGAAGCCCAATAGGAAATTTCTTCACCTC 3092
QY 36 -----GlnGlnValLeuCysSerSerValProLeuSer 46
Db 3093 AGTATTGAAGCTATTTTGGGAAGAATGTAAAGAGATGTGAAGGCTCT-----TCC 3143
QY 47 LysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIle 66
Db 3144 AAAAAAGTCACTCAGATT-----GAGGAAGTGAACACAAATTT 3182
```



```

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1028
; LENGTH: 6284
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-1028

```

## Alignment Scores:

```

Pred. No.: 1.72e-12 Length: 6284
Score: 220.50 Matches: 157
Percent Similarity: 36.2% Conservative: 118
Best Local Similarity: 20.7% Mismatches: 233
Query Match: 7.0% Indels: 251
DB: 3 Gaps: 36

```

US-10-644-084-2 (1-615) x US-09-949-016-1028 (1-6284)

```

QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
DB 3033 AGTAAGCAAGTTAAAGAATATCGAATTCAAGAACCCCAATAGGAAATAATCTTTCCACTCT 3092
QY 36 -----GlnGlnValLeuCysSerSerValProLeuSer 46
DB 3093 AGTATTGAAGCTATTGGGAAGAACTAAAGAGATTGTGAAGCCCTCT-----TCC 3143
QY 47 LysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIle 66
DB 3144 AAAAAAAGTCATCAGATT-----GAGGAACCTGGAACAACAATAAT 3182
QY 67 SerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGlnSerLys 86
DB 3183 GAAAAATTGCGAGGCAAGATAAA-----GGCTATAAGGATGAAAAACAATAAGACTAAAG 3236
QY 87 SerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeu 106
DB 3237 GAGAGGAGCATAAAAACCAAGATGACCTACTAAAA-----GAAAAAGAAACT 3284
QY 107 LeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSer-----ValGluThr 122
DB 3285 CTTATACAGCAGCTGAAGAAGAAATTGCAAGAAAAAATGTTACTCTTGATGTTCAATA 3344
QY 123 GlnAsnLeuLysLeuGly-----SerAspMetAspHisLeuGlnSerCysTyr 138
DB 3345 CAGCATGTAGTTGAAGGAAGAGAGCGCTTTTCAGAACTTACACAAGGTGTTACTTGCTAT 3404
QY 139 ---AlaLysLeuLysGluGln-----LeuGluThrSerArg----- 149
DB 3405 AAGGCAAAAATAAGGAACCTTGAACAATTTTAGAGACTCAGAAAGTTGAACGTAGTCAT 3464
QY 150 -----ArgGluMetIleGlyLeuGlnGluArgAsp 159
DB 3465 TCAGCAAGTTAGAACAAAGACATTTTGGAAAAAGGAATCTATC---ATCTTAAAGCTAGAA 3521
QY 160 ArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAsp--- 178
DB 3522 AGAAATTTGAAGGAATTTCAAGAACATCTTTCAGGATTCCTGTCAAAAAACCAACCAAGATTTA 3581
QY 178 ----- 178
DB 3582 AATGTAAAGGAACCTCAAGCTGAAGAGAAATCAACAGTTAAACAATAATTTGCAAGAT 3641

```

```

QY 179 -----GluValGln 181
DB 3642 ATGAAACATTACTTCAATTTAAAGAAGAAGAAGAAACCAACACGACGCAAGAAACAGAA 3701
QY 182 LysLeu---GlnAsnIleIleAlaSerArgAlaThrGlnTyrAsn-----HisAspVal 198
DB 3702 AAAATGAAAGAGGAACTCTCTCGAAGCTCTGCTCGTACCAGAACTCTGAAAGCAGATCTT 3761
QY 199 LysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsn 218
DB 3762 CAGAGGAAGGAAGAATATGCTGACCTGAAAGAAACTGACTGAT-----GCCAAA 3815
QY 219 LysLysAspLysAsnIleAlaMetAspValLeuLeuAsnTyrValGlyArgAlaAspGlyLys 238
DB 3816 AAGCAGATTAAAGCAAGTACAGAAAGAGGTA----- 3845
QY 239 ArgGlySerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIle 258
DB 3846 -----TCTGTAATGCGGTGATGAGGATAAATTTACTGAGGATT 3881
QY 259 LeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLys 278
DB 3882 AAAATTAATGAAGTGGAGAAAAAAGAAACCAAGTGTCTCAGCAATTAGATATGAAGCAG 3941
QY 279 LysValLeuGlnMetLysLysGluMetLysSerLeuLeuSerProGlnLysLysLys 298
DB 3942 CGAACCAATTCAGCAACTCAAGGAGCAGTTAAATAATCAGAAAGTGAAGCAAGCTATACAA 4001
QY 299 ProArgGluArgAla-----GluAspGlyThr 307
DB 4002 CAGTATGAGAGAGAGCTGCAAAAGATCTAAATGTTAAAGAGAAATAATTTGAAGACATGCGA 4061
QY 308 GlyThrValAlaIleSerAspIleGluAspSerGlyGluLeuSerArgAspSerVal 327
DB 4062 ATGACACTAGAA-----GAACAGGAACAACACTCAGTGAGAACAGGATCAAGTG 4109
QY 328 TrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTyr 347
DB 4110 CTTGAGGCTAAATTAGAGGAAGTT---GAAAGGCTGGCCACAGAAATTTGAAAAA---TGG 4163
QY 348 ArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGlu 367
DB 4164 AAGGAAAAATGCAATGATTTCGAAACCAAAACCAATCAAGGTCAATTAAGAACAATGAG 4223
QY 368 GlyLeuAsnGluGluAspValIleSerArg-----GlnAspHisGluGlnGlu 383
DB 4224 -----AACACACAGATGCTGCTGGAAAGCTCACTAATCTTCAAGATGAGTTACAGGAG 4277
QY 384 ThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGln 403
DB 4278 TCTGAACAG---AAATATAATGCTGATAGA---AAGAAATGCTTAGAAGAAAAAATGATG 4331
QY 404 LeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeuLeuArgAsp 423
DB 4332 CTTATCACTCAA----- 4343
QY 424 CysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTyrThrLeuPheLysGluGln 443
DB 4344 -----GCCAAAGAACGACAGAAATATACGAAT-----AAGAGATG 4379
QY 444 LysLysAsnPheGluArgGluArgSerPhe----- 454
DB 4380 AAAAAATATGCTGAGGACAGGAGCGTTTTTTTAAAGCAACAGAAATGAAATGGAATACTG 4439
QY 455 ThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg----- 472
DB 4440 ACAGCCAGCTGACAGAGAAAGATAGTGACCTTCAAAAGTGGCGGAGAGAAACGAGATCAA 4499
QY 473 -----AlaSerTyrValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGln 489
DB 4500 CTGTTGCGAGCTTTTAGAAATACAGCTAAAGCACTGATATCCAGTAATGTACAGAAAGAT 4559

```

Qy	490	Asn	-----SerGluAsnValLysLeuPheSerAlaPhe	500
Db	4560	AATGAAATTGAACAACATAAAAGGATCATATCAGAGACTTCTTAAAAATAGAAAACACAAATC	4619	
Qy	501	Ser	-----GlySerSerAspProAsp	507
Db	4620	ATGGATATCAAGCCCAACGCTATTAGTTACGAGATCCTGACAACTTCAAACTGACCT	4679	
Qy	507		-----	507
Db	4680	CTATCGAAGTTTTTGAAATTTCCAGAAAATAAATAGAGGATGGATCTGTAGTCCTTGAC	4739	
Qy	508		-----AsnLeuIle	510
Db	4740	TCTTGTGAAGTGTCAACAGAAAATGATCAAGCACTCGATTTCCAAACCTGAGTTAGAG	4799	
Qy	511	ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys	530	
Db	4800	ATTCAATTTTACACCTTTTACAGCCAAACAAATGGCAGTGAAGAAC	4853	
Qy	531	ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr	550	
Db	4854	ACCACACGATGACAGTTGAGATTTCCCAAGCT	4895	
Qy	551	HisSerCysValSerGluHisSerSerIle	567	
Db	4896	AAGAGTAAATGAATGGAGGAGGACTTCGTGAAATGTGAAAATAAGAAAGATGCTACACCC	4955	
Qy	568	GluGluSerLysPro	584	
Db	4956	AGAAGTAATTTGAAATTTCCATATTTAGATAGTAAATTTCTTCTGTCAAAAAGGACAA	5015	
Qy	585	SerValGlnSerArgProSerArgGluGlyCysTyrSerGlyCysSerAla	603	
Db	5016	AAGTGTCCCATACGTCTCATCATCAAGAAAACATATTTCTTACGAGTCAAGGA	5069	

US-10-644-084-2 (1-615) x US-09-643-597-117 (1-6921)

Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	1265	TTAGATGATTGGATC-----CAGCAGGTTGAAACTACTCAGAGAAAGATTACAGAA	1315
Qy	21	GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys	40
Db	1316	AATCAGCCTGANAATAGTAAACCCCTAGCACACAGTGTGAATCAACAGAAGATGCTGGTG	1375
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	1376	TCCGAAATAGAAATGAACACAGACAAATGGACGAG-----TGTCAAAATAT	1423
Qy	61	AsnIleGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	1424	GCAGAACAGTACTCAGCTCAGTGAAGGACTATTGAATTTACAAACAATGACTACCGGGCC	1483
Qy	81	LeuTyrGluGluSerLysSerLysGluAlaLysArg-----	92
Db	1484	ATGGTAGATTCAACAACAAATATCCAGTGAACACGCCGAAGATGCAGAGTTCAGCAGAT	1543
Qy	93	-----GluLeuAsnIleValAlaValLeuAsnCysMet	103
Db	1544	CTCATTATTCAAGAGTTTCATGGACCTTAAGACTCGATATACTGCCCTGGTCACTCTCATG	1603
Qy	104	AsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGln	123
Db	1604	ACACATATATATAAATTTGCTGGTGATTCATTGAGAGGCTGGAAGAGGAGGAGATTTAA	1663
Qy	124	AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysG	143
Db	1664	AGGTGAAGGAGACTTCTGAACATGGGCATATTTCAGATCTGCTTCAGCGTCAAGAAGCA	1723
Qy	143	uGln--LeuGluThrSerArg-----ArgGluMetIle	153
Db	1724	ACAGTGTCTGAGAAATAGCAAACTTACAGGAAGATAAAGTCAGTTGGAAGAAATGGTAGCT	1783
Qy	154	GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu	173
Db	1784	GAACTAAGAAACAAAGTCCCGAGTAGGAAGAAGACTCCCGAAGTTCAGGAGGCTGCA	1843
Qy	174	LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle	187
Db	1844	GAATAATCAATTGAGAAGCAGCAGAGAAAATGTGAAGAGATATCTCTGCAGAGAATAGG	1903
Qy	188	Ala--SerArgAlaThrGlnTyrAsnHisAsp-----ValLysArgLysGlu	202
Db	1904	GCTGAAGTGAACCAAGCAAGTACCAGGGAACTTGAACCACTTGTGAGAGAGAAGAA	1963
Qy	203	ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys	222
Db	1964	GCCGCTGAAAGAGAACTG--GAGCGGCTGAGGCAGCTCACCATAGAGGCCGAGGCTAA	2020
Qy	223	AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly	240
Db	2021	AGACGTGCCGTGAAGAGAACCTTCCTGAATTTTTTCGCAATCAGTGTGAGGAGAAACACCTTT	2080
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu	260
Db	2081	ACCAGACGAACACTGGAGATCATCTTAAAGAGAAAGAT-----TTAAGTCTC	2128
Qy	261	AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu-----	273
Db	2129	AATGATTTGGAG--CAACAAAAAATAAATTAATGGAAGAAATTAAGAAAGAGAGAC	2185
Qy	274	--AsnAlaGluLeuLysValLeuGlnMetLysLysGluMetIle-----	289
Db	2186	AATGAGGAGAGACTCTTGAAGCTGTATTAAGCAGATGGAAGAACCTTGCATTTCAGAA	2245
Qy	289	-----	289
Db	2246	CAGGTAGCAGAGAAACAGTTGAAAGAAAAAGCAGAAAATTAATTTGGAAGCAAGAGAAA	2305



```
QY 289 ----- 289
Db 2306 ATAACTGAAATTCAGTATACATAGAGAAAATGCCATTGCCAGTGTCCGATCACACAG 2365
QY 290 -----SerLeuLeuSerProGlnLysLysProArgGluArgAlaGlu 304
Db 2366 GCTACATCATGCAGGCGCAGTAAAGGCTCCAGCAGAGACATGACACAGCAGAAAGCAGAA 2425
QY 305 Asp-----GlyThrGlyThrValAlaIleSerAspIleGluAspSer 319
Db 2426 GAACTCAAAACAGCAGGTAGTAACTAAACAGCTGCCAATAGAAAGGCTGAACAGCATG 2485
QY 320 GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu 339
Db 2486 AGAGAGCTGACATATGAA-----CTTAATGCCCTCCAGCTTGAAAAA 2527
QY 340 ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn 359
Db 2528 AGTCATCTGAGAAAAGGCT---CGTTTGCTAAAGATAAACTAGATGAACAATAAT 2584
QY 360 -----GlnAlaSerLysValHisSer 366
Db 2585 ACACTCAGATGCTTAAAGTTGGAGCTGGAAAGGATCAGCGGAGAAAGGTTATTCT 2644
QY 367 GluGlyLeuAsnGlu-----GluAspVal 374
Db 2645 CAACAACCTCAGAGCTTGGTAGGCAATTGAATCAAAACACACAGGTAAAGCTGAAGAGCC 2704
QY 375 IleSer-----ArgGlnAspHisGluGlnGluThrGluLys 386
Db 2705 ATGCAGAAAGCTAGTGATCTCAAGAAAATAAAGCGCAATTATCAGTTAGTAATCT 2764
QY 387 LeuGluLeuGlu-----IleGluArgCysLysGlu 396
Db 2765 CTTAATCATGAAAAGGGAACTACAAGAGAGTAGACAGAAATCACAGGGCACATGCT 2824
QY 397 MetIleLysAlaGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAsp 416
Db 2825 GTAGCTGAGAGAATAATTCAGCATTTAAATTCACAAATTCATTCTTTTCAGATGAGAA 2884
QY 417 ThrThrSerLeuLeuArgAspCysThrLeuLeuGluLysGluArgLysGluGlu 436
Db 2885 GAAATTAGAAAGACTACAATCTGCCG-----AGAAATCAGATCATCTAAAGAACAA 2938
QY 437 TrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgSerPheThr 455
Db 2939 TTTGAGAAAAGCCATGAGCAGTTGCTTCAAAATATCAAAAGCTGAAAAGAAAAT---AAT 2995
QY 456 GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrp 475
Db 2996 GATAAAATCCAAAGGCTCAATGAAGAAATTGGAGAAA---AGTAATGAGTGTGCAGAGATG 3052
QY 476 ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys 495
Db 3053 CTAATAACAAAGAGTAGAGAGCTTACT-----AGCAGATAATGAACCAAAATTA 3103
QY 496 LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro 515
Db 3104 ATGATGCAGAGAATTCAGCAGCAATCA-----GAGAATATAGTTTAGAGAAACAAACT 3157
QY 516 ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer 532
Db 3158 ATCCAGCAAAAGATGTGAAGCACTGAAATTCAGGCAGATGGTTTAAAGATCAGCTACGC 3217
QY 533 LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis 551
Db 3218 ACCCAAATGAACACTTGCATTAACACAGACAAAACAGACGAGATTTTCAAAAGAAAAT 3277
QY 552 SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys 571
Db 3278 AAATGCCTAGAGAA-----GACCTGGCGAAAAGTCAAAATTTG 3316
```

```
QY 572 ProSerGluValAlaArgGluSerThrAspGlnLysThrSerValGlnSer 588
Db 3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACAGAACATTATCATCCAGAAT 3367
RESULT 11
US-09-480-884A-117
; Sequence 117, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A
; CURRENT FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 330
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 6921
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-480-884A-117
Alignment Scores:
Pred. No.: 4-86e-12 Length: 6921
Score: 217.00 Matches: 144
Percent Similarity: 38.4% Conservatives: 140
Best Local Similarity: 19.5% Mismatches: 268
Query Match: 6.9% Indels: 187
DB: 3 Gaps: 29
US-10-644-084-2 (1-615) x US-09-480-884A-117 (1-6921)
```

```
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20
Db 1265 TTAGATGATTGGATC-----CAGCAGGTTGAAACTACTCAGAGAAAAGATTCCAGAA 1315
QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 1316 AATCAGCTGAAATATGATAAAACCTAGCCACACAGATTGAATCAACAGAGATGCTGGTG 1375
QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 1376 TCCGAAATAGAAATGAAACAGACAGCAAAATGGACGAG-----TGTCAAAAATAT 1423
QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
Db 1424 GCAGAACAGTACTCAGCTACAGTGAAGGACTATGAATTACAAACAATGACCTACCGGCG 1483
QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArg----- 92
Db 1484 ATGGTAGATTCAACAAACAAATCTCCAGTGAACCCGCAAGAAATGCAGAGTTCAGCAGAT 1543
QY 93 -----GluLeuAsnIleValAlaValLeuAsnCysMet 103
Db 1544 CTCATTATTCAAGAGTTCATGACCTAAGGACTCATATATCTGCTGCTGCTCTCATG 1603
QY 104 AsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnSerValGlnThrGln 123
Db 1604 ACACAAATATATAAATTTGCTGGTGATTCAATTGAAGAGGCTGGAAGAGGAGAGATTA 1663
QY 124 AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGln 143
Db 1664 AGGTGAAGAGAGACTTCTGAAGATGGGCGCATATTACAGATCTGCTTCAAGCGTCAGAG 1723
QY 143 uGln-LeuGluThrSerArg-----ArgGluMetIle 153
Db 1724 ACAGTGCTTGAGATAGCAAACTTACAGAAAGAAATAGTAGTTCGAAAGAAATGTTAGCT 1783
```



Qy	1	MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer	20
Db	1265	TTAGATGTTGGTCTC-----CAGCAGGTTGAAACTACTCAGAGAAAGATTTCAGAA	1315
Qy	21	GlnTyThrSerGluThrLysMetSerProSerSerLeuTyThrSerGlnValLeuCys	40
Db	1316	AATCAGCCTGAAATAGTAAACCTTAGCCACACAGTGTGAATCAACAGAGATGCTGGT	1375
Qy	41	SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
Db	1376	TCCGAAATAGAAATGAACACAGACAGCAAAATGACGAG-----TGTCAAAATAAT	1423
Qy	61	AsnIleGluGlnSerIleSerTyLeuAspGlnGluLeuThrThrPheGlyPheProSer	80
Db	1424	GCAGAACAGTACTCAGCTACAGTGAAGGACTATCAATTTACAAACAATGACTACCGGCC	1483
Qy	81	LeuTyThrGluGluSerLysSerLysGluAlaLysArg-----	92
Db	1484	ATGGTAGATTCCAAACAAAATACTCCAGTGAACGCCGAAGATGTCAGAGTTTCAGCAGAT	1543
Qy	93	-----GluLeuAsnIleValAlaValLeuAsnCysMet	103
Db	1544	CTCATTTATTCAAGAGTTCATGGACTTAAGGACTCGATATAGTGCCTGGTCACTCTCATG	1603
Qy	104	AsnGluLeuLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGln	123
Db	1604	ACACAATATATTAATTTGCTGGTGATTCATTGAAGAGGCTGGAAGAGGAGGAGGATTTAA	1663
Qy	124	AsnLeuLysLeuGlySerAsp-MetAspHisLeuGlnSerCysTyzrAlaLysLysGlu	143
Db	1664	AGTGTAAAGGAGACTTCGAAACATGGGCGCATATTTCAGATCTGCTTCAGCCTCAGAAGCA	1723
Qy	143	uGln--LeuGluThrSerArg-----ArgGluMetIle	153
Db	1724	ACAGTCTTGAGATAGCAAACTTACAGGAAGATAAGTAGTGGTGAAGAAAGATGGTAGCT	1783
Qy	154	GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu	173
Db	1784	GAATAAAGAAACAAAAGTCCGAGTAGAGGAAGAACTTCCGAAGGTCAGGGAGGCTGCA	1843
Qy	174	LysAsnGluLysAspGluValGlnLys-----LeuGlnAsnIleIle	187
Db	1844	GAATAATGAATTGAGAAAGCAGCAGAGAAATGTAGAAGATATCTCTCTGCAGAGAATAAGG	1903
Qy	188	Ala---SerArgAlaThrGlnTyzrAsnHisAsp-----ValLysArgLysGlu	202
Db	1904	GCTCAAGTGNAGCCAAGCAGTACCCGAGGAACCTTGAACCAATTTGTGAGAGAGAGAA	1963
Qy	203	ArgGluTyzrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys	222
Db	1964	CCCCTGAAGAGAACTG--GAGCGGCTGAGGCAGCTCACCATTAGAGGCCGAGGCTAAA	2020
Qy	223	AsnIleAlaMetAsp-----ValLeuAsnTyzrValGlyArgAlaLysGlyArgGly	240
Db	2021	AGAGCTCCGCTGGAAGAGAACTCTCTGAATTTTCGCAATCAGTTGGAGGAAAAACACCTTT	2080
Qy	241	SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyzrLysIleLeuLeu	260
Db	2081	ACCAGACGACCTGGAAGATCATCTTTAAAGAAAAGAT-----TTAAGTCTC	2128
Qy	261	AsnAspTyzrGluTyzrArgGlnLysGlnIleLeuMetGlu-----	273
Db	2129	AATGATTTGGAG--CAACAAAAATAATTAATGGAAGAAATTAAGAAAGAGAGAGAC	2185
Qy	274	---AsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle-----	289
Db	2186	AATGAGGAAGAACTCTTGAAGCTGATAAAGCAGATGGAAGAAAGACCTTGCTTCAGAAA	2245
Qy	289	-----	289
Db	2246	CAGTACGAGAGAAACAGTTGAAAGAAAACGAGAAATTAATTTGGNAGCAAGAGAAA	2305
Qy	289	-----	289

2306	DB	ATAACTGAAATTCAGTATACATGATAGAGAAAAATGCATTGCCAGTGTCTCCGATCCACACAG	2365
230	QY	-----SerLeuLeuSerProGlnLysLysLysProArgGluuArgAlaGlu	304
2366	DB	CTACATCATCGCGCAGTAACGGGTCTCCAGCAAGAACATGACAGACAGAAAGCAGAA	2425
305	QY	-----GlyThrGlyThrValAlaIleSerAspIleGluuAspAspSer	319
2426	DB	GAACCTCAAACACAGCAGGTAGATGAATTAACAGCTGCCAATAGAAGGCTGAACACAGACATG	2485
320	QY	GlyGluLeuSerArgAspSerValTrrpGlyLeuSerCysAspThrValArgGluGlnLeu	339
2486	DB	ACAGAGCTGCACATATGAA-----CTTAATGCCCTCCAGCTTTGAAAAA	2527
340	QY	ThrAsnSerIleArgLysGlnTrrpArgIleLeuLysSerHisValGluLysLeuAspAsn	359
2528	DB	ACGTCACTCTGAGGAAAAAGGCT---CGTTTGTCTAAAGATATAACTAGATGAACCAATAAT	2584
360	QY	-----GlnAlaSerLysValHisSer	366
2585	DB	ACACTCAGATGCCCTTAAGTTGGAGCTGGAAGGAAGGATCAGGCGGAGAAAGGGTATTCT	2644
367	QY	GluGlyLeuAsnGlu-----GluAspVal	374
2645	DB	CAACAACCTCAGAGAGCTGTGTAGCAATTCGAATCAAAACCACAGGTAAGCTGAAGACCC	2704
375	QY	IleSer-----ArgIlnAspHisGluGlnGluThrGluLys	386
2705	DB	ATGCAAGAAGCTAGTGTATCTCAAGAAAAATAAAGCGCAATTATCAGTTTGAATTTAGAATCT	2764
387	QY	LeuGluLeuGlu-----IleGluArgCysLysGlu	396
2765	DB	CTTAATCATGAAAAAGGAACTACAAGAGAAAGTAGACAGAAATCAACAGGGCAGCATGCT	2824
397	QY	MetIleLysAlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrCysAspAspAsp	416
2825	DB	GTACTCAGAGAAGATATTCAGCATTTAAATTCANAATTCATCTCTTTTCAGATGAGAAA	2884
417	QY	ThrThrSerLeuLeuArgAspCysTyrrLeuLeuGluGluLysGluArgLeuLysGluGlu	436
2885	DB	GAATTAGANAAGACTACAAATCTGCCAG-----AGAAAATCAGATCATCTATAAAGAACAA	2938
437	QY	TrrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgArgSerPheThr	455
2939	DB	TTTGAGAAAAAGCCATGACGAGTGTCTTCAAAATATCAAAAGCTCAAAAGAAAAAT---AAT	2995
456	QY	GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrrp	475
2996	DB	GATAAAATCCNAGGCTCAATGAAGAATTCGAGAAA---AGTAATGAGTGTGAGAGATG	3052
476	QY	ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys	495
3053	DB	CTAAAAAATAAAGTAGAGGAGCTTACT-----AGGCAGAAATAATGAACCAAAATTA	3103
496	QY	LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro	515
3104	DB	ATGATGCAGAGAATTCAGGCAGAAATCA-----GAGAATATAGTTTTAGAGAAACAAACT	3157
516	QY	ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer	532
3158	DB	ATCCAGCAAGATCTGAAGCACTGMAAATTCAGGCAGATGGTTTTAAAGATCAGCTACGC	3217
533	QY	LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis	551
3218	DB	AGCAAAATGAACACTTTCGATAAACACAGACAAAAACAGACAGCAGGATTTTCAAGAAAAAT	3277
552	QY	SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys	571
3278	DB	AAATGCCCTAGAAGAA-----GACTGCGGAAAAAGTCAAAATTTG	3316
572	QY	ProSerGluValAlaArgGluSerThrAspGlnLysTrrpSerValGlnSer	588

[illegible]





Db 1904 GCTGAAAGTGAAGCAGCAGTACCGCAGGAACTTGAACCATTTGTGAGAGAGAGGAA 1963  
Qy 203 ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLys 222  
Db 1964 GCCCGCTGAAAGAGAACTG---GAGCGGCTGAGGCAGCTCACCATAGAGCGGAGGTAAA 2020  
Qy 223 AsnIleAlaMetAsp-----ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db 2021 AGAGCTGCCGTGAAGAGAACTCCTGGAATTTTCGCNAATCAGTTGGAGGAAAAACACCTTT 2080  
Qy 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
Db 2081 ACCAGAGCAACACTGGAAGATCATCTTAAAGAAAGAT-----TTAAGTCCTC 2128  
Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGlu----- 273  
Db 2129 AATGATTTGGAG---CAACAAAAAAATAAATTAAATGAAGAAATTAAGAGAAAGAGAGAC 2185  
Qy 274 ---AsnAlaGluLeuLysValLeuGlnGluMetLysLysGluMetIle----- 289  
Db 2186 AATGAGGAAGAACTCTTGAAGCTGATAAGCAGATGGAAGAAAGACCTTGCAATTCAGAAA 2245  
Qy 289 ----- 289  
Db 2246 CAGGTAGCAGAGAAACAGTTGAAAGAAAGCAGAAATTTGAATTGGAAGCAAGAGAAA 2305  
Qy 289 ----- 289  
Db 2306 ATAACCTGAAATTCAGTATACATGTAGAGAAATGTCATTGCCAGTGTGTCGATCACACAG 2365  
Qy 290 -----SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGlu 304  
Db 2366 GCTACATCATGTCAGGCGAGTAACGGGTCTCCAGCAAGAACATGCACAAGCAGAAAGCAGAA 2425  
Qy 305 Asp-----GlyThrGlyThrValAlaIleSerAspIleGluAspSer 319  
Db 2426 GAACCTCAACAGCAGGTAGTAGTAACACAGCTGCCAATAGAAAGGCTGAACAGACATG 2485  
Qy 320 GlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeu 339  
Db 2486 AGAGAGCTGCATATGAA-----CTTAATGCCCTCCAGCTTGAAAAA 2527  
Qy 340 ThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn 359  
Db 2528 ACGTCACTCAGGAAAAAGGCT---CGTTTGCTTAAAGATATAAAGCTAGATGAAACAAATAAT 2584  
Qy 360 -----GlnAlaSerLysValHisSer 366  
Db 2585 ACACCTCAGATGCTTAAGTTGGAGCTGGAAGAGGATCAGCGCGAGAAAGGGTATTCT 2644  
Qy 367 GluGlyLeuAsnGlu-----GluAspVal 374  
Db 2645 CAACAACTCAGAGAGCTTGGTAGGCAATTGMAATCAAAACACAGGTAAAGCTGAAGAAGCC 2704  
Qy 375 IleSer-----ArgGlnAspHisGluGlnGluThrGluLys 386  
Db 2705 ATGCAAGAAAGCTAGTGATCTCAAGAAAAATAAAGCGCAATTATCAGTTAGAAATTAGAATCT 2764  
Qy 387 LeuGluLeuGlu-----IleGluArgCysLysGlu 396  
Db 2765 CTTAATCATGMAAAGGGAACACTACAAGAGAGATAGACAGAAATCAAGGGCACATGCT 2824  
Qy 397 MetIleLysAlaGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAsp 416  
Db 2825 GTAGCTGAGAAGAATATTTCAGCATTTAAATTCACAAATTCATTCTTTTCGAGATGAGAAA 2884  
Qy 417 ThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArgLysGluGlu 436  
Db 2885 GAATTAGAAAGACTACAAAATCTGCCAG-----AGAAAAATCAGATCATCTTAAAGAACAA 2938  
Qy 437 TrpThrLeuPheLysGluGln---LysLysAsnPheGluArgGluArgSerPheThr 455  
Db 2939 TTTGAGAAAAGCCATGAGCAGTTGCTTTCAAAATATCAAAAGCTGAAAAAGAAAAT---AAT 2995

Qy 456 GluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrp 475  
Db 2996 GATAAAATCCAAAGGCTCAATGAAGAATTTGGAGAAA---AGTAATGAGTGTGCAGAGATG 3052  
Qy 476 ValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLys 495  
Db 3053 CTAAAAACAAAAAGTAGAGAGCTTACT-----AGGCAGAAATTAATGAACCAAAATTA 3103  
Qy 496 LeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgPro 515  
Db 3104 ATGATCCACAGAAATTCAGGCAGAAATCA-----GAGAAATATAGTTTTAGAGAAACAAACT 3157  
Qy 516 ArgGlnLysLysLeuHisSerVal-----AlaAsnGlyValProAlaCysThrSer 532  
Db 3158 ATCCAGCAAAAGATGTGAAGCACTGMAAATTCAGGCAGATGGTTTTAAAGATCAGCTACGC 3217  
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThr---SerAspPheArgGlnThrHis 551  
Db 3218 AGCACAATGAACACTTGCATAACACAGACAAAAACAGACAGAGATTTTCAAGAAAAAAT 3277  
Qy 552 SerCysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLys 571  
Db 3278 AAATGCTAGAGAA-----GACCTGGCGAAAAAGTCAAAATTTG 3316  
Qy 572 ProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSer 588  
Db 3317 GTAAGTGAATTTAAGCAAAAGTGTGACCAACACAGACATTATCATCCAGAAT 3367

Search completed: June 14, 2006, 01:52:55  
Job time : 434 secs

**THIS PAGE LEFT BLANK**



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.  
OM nucleic - nucleic search, using sw model  
Run on: June 13, 2006, 22:27:49 ; Search time 2843 Seconds  
(without alignments)  
11635.003 Million cell updates/sec  
Title: US-10-644-084-1  
Perfect score: 2692  
Sequence: 1 cgtaggagagtgacaggagc.....ttgattatggagtaatgggg 2692  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 18992170 seqs, 6143817638 residues  
Total number of hits satisfying chosen parameters: 37784340  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA Main:\*

1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq.\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq.\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1258.8	46.8	2272	7	US-10-104-047-1497
2	1258.8	46.8	2272	16	US-11-072-512-1497
3	399.8	14.9	161874	15	US-11-121-086-75
4	305.8	11.4	403	10	US-10-779-543-12377
5	130.6	4.9	513	7	US-10-029-386-5594
6	123.6	4.6	178	7	US-10-029-386-19350
7	61.2	2.3	30191	11	US-10-330-773-631
8	60.4	2.2	621	10	US-10-450-763-26492
9	57.4	2.1	3750	8	US-10-437-963-34854
10	55.8	2.1	2093	7	US-10-104-047-1666
11	55.8	2.1	2093	16	US-11-072-512-1666
12	54.6	2.0	13407	10	US-10-737-318-27
13	54	2.0	1199	8	US-10-425-114-10108
14	54	2.0	1567	8	US-10-424-599-27277
15	53.2	2.0	154817	7	US-10-085-117-334
16	52.4	1.9	1062	9	US-10-767-795-5235
17	50.8	1.9	556	3	US-09-732-627A-536

c	18	50.8	1.9	193853	6	US-10-087-192-1663	Sequence 1663, Ap
	19	49.8	1.8	12025	7	US-10-311-455-1271	Sequence 1271, Ap
	20	49.8	1.8	19380	8	US-10-221-613-389	Sequence 389, App
	21	49.4	1.8	1640	10	US-10-750-185-37160	Sequence 37160, A
	22	49.4	1.8	1640	10	US-10-750-623-37160	Sequence 37160, A
	23	49.4	1.8	6283	8	US-10-221-613-313	Sequence 313, Appl
	24	49.4	1.8	52138	13	US-11-166-990-75	Sequence 75, Appl
c	25	49.4	1.8	92726	3	US-09-997-722-193	Sequence 193, App
	26	49.2	1.8	633	8	US-10-338-110-121	Sequence 121, App
	27	48.8	1.8	1248	8	US-10-282-122A-40644	Sequence 40644, A
	28	48.8	1.8	3773	3	US-09-925-302-47	Sequence 47, Appl
	29	48.8	1.8	3773	3	US-09-925-302-47	Sequence 47, Appl
	30	48.8	1.8	4041	7	US-10-264-049-543	Sequence 543, App
	31	48.8	1.8	5883	10	US-10-356-157-859	Sequence 859, App
	32	48.8	1.8	6354	6	US-10-084-817-158	Sequence 158, App
	33	48.8	1.8	6432	10	US-10-450-763-1707	Sequence 1707, Ap
	34	48.8	1.8	7274	10	US-10-696-909A-48	Sequence 48, Appl
	35	48.8	1.8	7396	7	US-10-028-248A-35	Sequence 35, Appl
	36	48.8	1.8	7396	8	US-10-107-782-35	Sequence 35, Appl
	37	48.8	1.8	7474	15	US-11-069-834-49	Sequence 49, Appl
c	38	48.8	1.8	7736	10	US-10-450-763-1714	Sequence 1714, Ap
	39	48.8	1.8	24801	10	US-10-684-042-18	Sequence 18, Appl
c	40	48.8	1.8	25301	8	US-10-671-007-3	Sequence 3, Appli
	41	48.4	1.8	224112	8	US-10-367-054-80	Sequence 80, Appl
	42	48.2	1.8	5728	7	US-10-311-455-73	Sequence 73, Appl
	43	48.2	1.8	3673778	7	US-10-312-841-2	Sequence 2, Appli
	44	48	1.8	835	8	US-10-282-122A-35806	Sequence 35806, A
	45	48	1.8	37265	6	US-10-087-192-49	Sequence 49, Appl

ALIGNMENTS

RESULT 1  
US-10-104-047-1497  
; Sequence 1497, Application US/10104047  
; Publication No. US00030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1497  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1497

Query Match	45.8%	Score 1258.8;	DB 7;	Length 2272;
Best Local Similarity	77.2%;	Pred. No. 0;		
Matches 1651;	Conservative	0;	Mismatches 337;	Indels 150; Gaps 4;
Qy	1	CGTAGGAGAGTGACAGAGCTGTTGTAAAGCGTCGACACCTGAGCGCCTCTCAGGAT 60		
Db	145	CATAGAAGTGACAGAGCTAGACTAAATGTTTAACTGCTGAACCTAGTCTCAGGAT 204		
Qy	61	CTTGCTCTGGAACCTCTCTATGGAGATTGGATGACTGTGACAGATCCAGTCTGTGTAC 120		
Db	205	CTTGCTCTAGAGATTGCTATGGAGATTGGATGACTGTGTACAGATC-----C 252		
Qy	121	AGAAAAAATAATCTCTCAATATACCTCAGAAACAAAGATGCTCGTCCAGTTGTA 180		
Db	253	AGAAAGCAAACTATCTCTCAATATACCTCAGAAACAAAGATGCTCTCAAGTTTATA 312		
Qy	181	CTCCACCAAGTCTGTGCTCTTCAGTACCTTTATCAAAAAACGTGATGTTTTCGG 240		
Db	313	CTCACCAAGTGTCTGTCTTCAATACCTTTTATCAAAAAATGTGCACAGTTTTTCAG 372		









```
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 631
; LENGTH: 30191
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(30191)
; OTHER INFORMATION: n = A,T,C or G
US-10-330-773-631

Query Match      2.3%; Score 61.2; DB 11; Length 30191;
Best Local Similarity 43.7%; Pred. No. 0.0002;
Matches 413; Conservative 0; Mismatches 528; Indels 5; Gaps 3;

Qy 366 TAGTCGCTGTTCTGAACTGTATGAACGAGCTGCTCGTCTTTCAGCGGAAGAACCTGCTGG 425
Db 14557 TTGAAGCTCGCAGAGCAGGAGCAGGAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14498

Qy 426 CCCAGGAGAGCTGGAGACACAGAACTTGAAGCTGGGCGAGTGACATGACCACCTGCAGA 485
Db 14497 AGCAGGAGAGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14438

Qy 486 GCTGCTACGCCAAACTTAAGGAGCAGTTGGAAGAACTCCAGCGCGGAGATGATCGGC--- 542
Db 14437 AGGAGGAAGACGAGGAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGG 14378

Qy 543 TTCAAGAGAGAGCAGGAGCTGCAGTGCAAGAAACAGGAGTTTGCATCAGCTCCTGAAGA 602
Db 14377 AGCAGGAGCAGGAGGAAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14318

Qy 603 ATGAGAAAGATGAGGTACAAAAATTAACAAATATCATGACCGCGGGCTACTCAGTATA 662
Db 14317 AGGAGGAGCAGGAGGAAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14258

Qy 663 ATCATGATGTGAAGAGGAAGAGCGTGTAATATAATAGCTAAAGAGCGCTGCATCAGC 722
Db 14257 AGGAGCAGGAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14198

Qy 723 TCGTTATGAACAAGAGGATAAACATAGCCATGGATGTTTTAAATTTATGTGGGTCCAG 782
Db 14197 AGGAGGAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14138

Qy 783 CTGATGGCAACGAGGCTCATGGAGGACTGACAAAACAGAACCCAGGAAATGAAGATGAGA 842
Db 14137 AGGAGCAGGAGGAAGACAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 14078

Qy 843 TGTCAAAATTTCTGTTGAATGATTATGATACCGCAGAACAGATCCTGATGAGAAACG 902
Db 14077 AGCAGGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 14018

Qy 903 CGGAGCTGAAGAAGTCTCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGTCCTC 962
Db 14017 AGGAGCAGGAGCTGGAGCAGGAGTAGGATAGGAGTAGGAGGAGGAGGAGGAGGAGGAGC 13958

Qy 963 AGAAGNAGAACCCAGGAAAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1022
Db 13957 AGGAGCTGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13898

Qy 1023 TAGAAGATGACTCTGGGGAACCTG-AGCAGACACAGCGTGTGGGGCTTTCTGTGACACT 1081
Db 13897 AAGAGCAGGAGCTGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 13838

Qy 1082 GTGAGAGAGCAGCTGACAAAACAGGATCAGGAAACAGTGGAGAAATTTTGAAGTCAATGTA 1141
Db 1141
```

---

```
Db 13837 AGGAGCAGGAGGAGGAAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 13779
Qy 1142 GAAAACTCGATAACCAAGCTTCGAAGGTACACTCAGAGGCGCTTAATGAGGAGCGTC 1201
Db 13778 CAGGAGGAAGCAGGAGGCTGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAG 13719
Qy 1202 ATCTCACGACAAACCATGAGCAAGAGACTTGAGAAACTGGAGCTGGAGATTGACGGTGT 1261
Db 13718 CAGGAGCAGGAGGAGGAAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13659
Qy 1262 AAAGAGATGATCAAGGCTCAGCAGCAGCTCTTACAGCAGCAGCTGG 1307
Db 13658 CTGGAGCTGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13613

RESULT 8
US-10-450-763-26492
; Sequence 26492, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 26492
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (269)..(595)
; OTHER INFORMATION: 35% homologous to Plasmodium falciparum glutamic acid-rich
; OTHER INFORMATION: protein, accession number AF251290.Smith-Waterman Score=193.
US-10-450-763-26492

Query Match      2.2%; Score 60.4; DB 10; Length 621;
Best Local Similarity 48.3%; Pred. No. 2.7e-05;
Matches 228; Conservative 0; Mismatches 241; Indels 3; Gaps 2;

Qy 503 AAGGAGCAGTTGGAACGTCCTCCAGCGGAGATGATCGGGCTTCAAGAGAGAGCAGGCAG 562
Db 103 AAGAGGAAGAGGAGGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 162

Qy 563 CTGCAGTGCAGAACACAGGAGTTTGCATCAGCTCCTGAGAGTGAAGAGATGAGGTACAA 622
Db 163 AAGAAGAGAGGAGGAGGAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 222

Qy 623 AAATTACAAAATATCATAGCCCGGCTACTCAGTATATATCATGATGTGAAGAGGAAG 682
Db 223 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 282

Qy 683 GAGCGTGAATATATAAGCTTAAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAGAT 742
Db 283 AAGGAGAG--AGAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 340

Qy 743 AAAAACTAGCCATGGATGTTTTAAATATATGTGGTTCGAGCTGATGCGAACAGGCTCA 802
Db 341 AAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 400

Qy 803 TGGAGGAGCTGACAAAACAGAGAACCCAGGAATGAAGATGAGATGTACAAAAATTTCTGTGAAT 862
Db 401 GGAGAGAAATCTAGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 460

Qy 863 GATTATGAGTACCCCCAGAGAGCAG-ATCCTGATGGAGAACCCGGAGCTGAAGAAGGTCCT 921
```



```

; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1666
; LENGTH: 2093
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-072-512-1666

```

```

; Sequence 27, Application US/10737318
; Publication No. US20050202442A1
; GENERAL INFORMATION:
;
; APPLICANT: Morris, David W.
; APPLICANT: Malandro, Marc S.
; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
; FILE REFERENCE: 529452002800
; CURRENT APPLICATION NUMBER: US/10/737,318
; CURRENT FILING DATE: 2003-12-15
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 13407
; TYPE: DNA
; ORGANISM: Mus musculus
; US-10-737-318-27

```



	Matches	190;	Conservative	0;	Mismatches	195;	Indels	9;	Gaps	1;
Qy	589	TCAGCTCCTCGAAGAAATGAGAAGATGAGGTACAAAATTACAAAATATCATAGCCAGCCG	648							
Db	142	TCGAAGCTGCAACAGGAGGCGATGAATTCACAGAAATGGTTATTGGTAAATCAGCAAGT	201							
Qy	649	GGCTACTCAGTATAATCATGATGTGMAAGAGGAGCGGTCAATATATAATTAAGCTTAAGGA	708							
Db	202	AAAGACTCAACAATATGCATGATGATGAAGAAGAGGAAAGAAATACATAAAGTTTGCAGGA	261							
Qy	709	CGCGCTGCATCAGCTCGTTATGAACAAGAAGGATATAAAACATAGCCATGGATGTTTTAAA	768							
Db	262	GAGGCTAAACCAAGTGTGATGGAGAGAGAAGAGAGTCTAGGTCAGGCATGGAGATAAT	321							
Qy	769	TTATGTGGGTGCGAGCTGATGCCAAACGAGGCTCATGGAGGACTGACAAAACAGAGCCAG	828							
Db	322	GAATTTTGCTTCAGAGAAGGAGGAGCAACGTGGAAACATGGAATGGGAAGAAAGCTGACA	380							
Qy	829	GAATGAAGATGAGATGTACAAAAATTCGTTGCAATGATTATGAGTACCGCCAGAGCAGAT	888							
Db	381	-----ATGATTTTTATAAAAAGATTGTGGATGCTTATGAGTCAAAAATATCAAGAACT	432							
Qy	889	CCTGATGGAGAACCGGAGCTGGAAGAAGGTCTCCAGCAGATTAAGAAGGAGATGATCTC	948							
Db	433	GATGGCAGAGAAATGCTGATTTAAGAGCATATTTCAGATCAATGCAAGGTGATATCGGTGA	492							
Qy	949	TCTCCTGTCTCCTCAGAGAAGAAAGCCAGGGAA	982							
Db	493	TTTCTTAAATGCTCCTAATGGACTACCAAAAGCAA	526							

```

RESULT 14
US-10-424-599-27277
; Sequence 27277, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 27277
; LENGTH: 1567
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124632C.1
US-10-424-599-27277

```

	Query Match	2.0%	Score 54;	DB 8;	Length 1567;
	Best Local Similarity	48.2%;	Pred. No. 0.003;		
	Matches 190;	Conservative 0;	Mismatches 195;	Indels 9;	Gaps 1;
Qy	589	TCAGCTCCTGAAGAAGTACAGAAAGATGAGGTACAAAAATTACAAATATCATAGCAGCCG	648		
Db	508	TGAGAGCTGCACAGAGAGGGATGAAATCCAGAGAAATGGTATTGGTAAATCAGCAAGT	567		
Qy	649	GGCTACTCAGTATAATCATGATGGAAGAGAGGAGCGGTGAATATAATTAAGCTTAAGGA	708		
Db	568	AAAGACTCAACAATATGTCATGAGATGGAAGAGAGGAAAAAGAAATACATAAAGTTGCAGGA	627		
Qy	709	CGCGCTGCATCAGCTCGTATTGAACAAGAAGGATAAAAACATAGCCCATGCATGTTTTAAA	768		
Db	628	GAGGCTAAACCAAGTGTGTATGGNAGAGAGAAAGAGTCTTAGGTCAGGCATGAGATTAAT	687		
Qy	769	TTATGTGGGTGAGCTGATGGCAACAGAGGCTCATGGAGAGCTGACAAAAACAGAAGCCAG	828		
Db	688	GAATTTTGCTTTCAAGAAGAGGGAGGCACGTGGAAACATGGAATGGGAAGAAAGCTGCACA	746		

Qy	829	GAATGAAGATGAGATGTA	CAAAAATTCGTTGAATATTAGTACCGCCAGAA	GAGAT	888
Db	747	-----ATGATTTTAT	AAAAAGATTGCGATGCTTATGACTCAAAAAT	ATCAAGAACT	798
Qy	889	CCTGATGGGAACCGGAGCT	GAAAGAAAGTCTCTCCAGCAGATGAAGAAG	GAGATGATCTTC	948
Db	799	GATGGCAGAGAATGCTGAT	TAAAGACATTTATGAGATCAATCAGGTG	GATATGCGTGA	858
Qy	949	TCTCCTGTCTCTCAG	AAGAAAGAGCCCAAGAA		982
Db	859	TTTCTTAAATGCTCCT	TAATGGATACCAAGCAA		892

## RESULT 15

```

US-10-085-117-334/C
; Sequence 334, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 154817
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(154817)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-334

```

[illegible]

Search completed: June 14, 2006, 02:09:22  
Job time : 2851 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 13, 2006, 23:04:18 ; Search time 71 Seconds  
(without alignments)  
4812.369 Million cell

**Title:** US-10-644-084-1

**Perfect score:**

Sequence: 1 cgtaggagagtgcaggagc.....ttgattatggagtaatgggg 2692

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

0  
2  
4  
3  
2  
0  
2<sub>4</sub>  
3  
0  
—  
9  
9  
9  
1  
  
2<sub>4</sub>  
9  
2<sub>4</sub>  
1  
1

Searched: 253354 seqs, 63461778 residues

Total number of hits satisfying chosen parameters: 506708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Listing first 45 summaries

Database : Published Applications NA New.\*

Database	Finished Applications	NA	New	1	2	3	4	5	6	7	8
	1	/EMC_Celerra_S1DS3	ptodata/1/pubna/US09	NEW	PUB	seq:*					
	2	/EMC_Celerra_S1DS3	ptodata/1/pubna/US06	NEW	PUB	seq:*					
	3	/EMC_Celerra_S1DS3	ptodata/1/pubna/US07	NEW	PUB	seq:*					
	4	/EMC_Celerra_S1DS3	ptodata/1/pubna/US08	NEW	PUB	seq:*					
	5	/EMC_Celerra_S1DS3	ptodata/1/pubna/PCT	NEW	PUB	seq:*					
	6	/EMC_Celerra_S1DS3	ptodata/1/pubna/US10	NEW	PUB	seq:*					
	7	/EMC_Celerra_S1DS3	ptodata/1/pubna/US11	NEW	PUB	seq:*					
	8	/EMC_Celerra_S1DS3	ptodata/1/pubna/US60	NEW	PUB	seq:*					

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	52.2	1.9	1497	6	US-10-953-349-6159	Sequence 6159, Ap
2	45.4	1.7	2479	7	US-11-293-697-1966	Sequence 1966, Ap
3	42.6	1.6	759	7	US-11-217-529-1065	Sequence 1065, Ap
4	40.2	1.5	1799	7	US-11-293-697-1885	Sequence 1885, Ap
5	38.8	1.4	1307	7	US-10-953-349-6314	Sequence 6314, Ap
6	38.8	1.4	3207	6	US-11-217-529-80932	Sequence 80932, A
C 7	38.8	1.4	5849	7	US-11-270-287-1	Sequence 1, Appli
C 8	38.4	1.4	11978	7	US-11-257-851A-63	Sequence 63, Appl
9	38.2	1.4	1046	6	US-10-953-349-23270	Sequence 23270, A
10	38	1.4	1438	7	US-11-301-554-1028	Sequence 1028, Ap
11	37.4	1.4	1434	7	US-11-217-529-80879	Sequence 80879, A
12	37.2	1.4	8998	6	US-10-953-349-36772	Sequence 36772, A
13	36.6	1.4	3495	7	US-11-217-529-5994	Sequence 5994, Ap
C 14	36.4	1.4	138941	6	US-10-489-730-10	GENERAL INFORMATI
C 15	36.4	1.4	1137	7	US-11-217-529-79782	Sequence 79782, A
C 16	36.4	1.4	1998	6	US-10-471-571A-3859	Sequence 3859, Ap
17	36.4	1.4	2782	6	US-10-473-173-29	Sequence 29, Appl
C 18	36.2	1.3	591	7	US-11-217-529-152	Sequence 152, App
19	36.2	1.3	1854	6	US-10-471-571A-441	Sequence 441, App
20	36	1.3	1158	7	US-11-217-529-1020	Sequence 1020, App
21	36	1.3	3161	7	US-11-293-697-1992	Sequence 1992, Ap
22	35.8	1.3	363	7	US-11-217-529-173450	Sequence 173450,
C 23	35.8	1.3	795	7	US-11-217-529-77171	Sequence 77171, A
C 24	35.8	1.3	1287	6	US-10-471-571A-1659	Sequence 1659, Ap
25	35.4	1.3	534	7	US-11-217-529-82693	Sequence 82693, A

```
Db      870 GCGTCTCTTCTTGAATGCT 888
      ||| ||| ||| ||| |||
RESULT 2
US-11-293-697-1966
; Sequence 1966, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1966
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1966
Query Match      1.7%; Score 45.4; DB 7; Length 2479;
Best Local Similarity 44.7%; Pred. No. 0.0099;
Matches 264; Conservative 0; Mismatches 321; Indels 6; Gaps 2;
Cy 403 GCTTCAGCGGAAGAACCTGCTGGCCAGAGGAGCGTGGAGACACAGAACTTTGAAGCTGGG 462
Db 1270 GCAGGAGGAGGAAGATACCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1329
Qy 463 CAGTGACATGACCACTGACAGCTGTACGCCAAACTTAAGAGCAGTGTGGAACGTC 522
Db 1330 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1389
Qy 523 CAGGCGGAGATGATCGGGCTTCAAGAGAGACAGCAGCTGCAGTGCAAGACAGGAG 582
Db 1390 GTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1449
Qy 583 TTTGCATCAGCTCTCTGAAGATGAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGC 642
Db 1450 GGAGCAGGAGGAGATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1509
Qy 643 CAGCGGGCTACTCAGTATATCATATGATGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 702
Db 1510 GCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1569
Qy 703 AAAGGAGCGCTGCATCAGCTCGTTTATGAACAAGAGGATAAAAACATAGCCATGGATGT 762
Db 1570 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1625
Qy 763 TTTAAATTTATGTGGTTCAGCTGATGGCAACGAGGCTCATGGAGGACTGCAAAAACAGA 822
Db 1626 TGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1685
Qy 823 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 882
Db 1686 AGATACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1743
Qy 883 GCAGATCCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 942
Db 1744 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1803
Qy 943 GATCTCTCTCTGCTCTCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 993
Db 1804 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1854
RESULT 3
US-11-217-529-1065
; Sequence 1065, Application US/11217529
; Publication No. US20060099612A1
```

```
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1065
; LENGTH: 759
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-1065
Query Match      1.8%; Score 42.6; DB 7; Length 759;
Best Local Similarity 56.9%; Pred. No. 0.029;
Matches 78; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 802 ATCGAGGACTGACAAACAGAACCCAGGATGAAGATGAGATGTACAAAATTTCTTTGAA 861
Db 453 AGAAAGAGGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 512
Qy 862 TGATTATGATTACCGCCAGAACGAGATCTCTGATGGAGAACCGCGAGCTGAAGAGGTCT 921
Db 513 GGACAAAAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 572
Qy 922 CCAGCAGATGAAGAAGG 938
Db 573 GAAAAAGGAGAAAAAGG 589
RESULT 4
US-11-293-697-1885
; Sequence 1885, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1885
; LENGTH: 1799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1885
Query Match      1.5%; Score 40.2; DB 7; Length 1799;
Best Local Similarity 51.5%; Pred. No. 0.22;
Matches 121; Conservative 0; Mismatches 108; Indels 6; Gaps 1;
Qy 388 GAACGAGCTGCTCGTCTTCAGCGGAGAACCTCTGGCCAGAGAGCGTGGAGACACA 447
Db 1187 GGAGAGAGATCGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1246
Qy 448 GAACTTGAAGCTGGCAGTGCATGTGACCACTCTGCAGAGCTGTCTACCCCAAACCTTAAG 507
Db 1247 GGAGGAGAGATCGCGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1306
Qy 508 GCAGTTGGAAAACGCTCCAGCGG-----GGAGATGATCGGCTTCAAGAGAGAGACAGCA 561
Db 1307 GCAGGAGGAGAGATGCGGGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1366
```

Qy 562 GCTGCAAGACAGGAGTTTGCATCAGCTCCTGAAGATCAGAAAGATGAG 616  
Db 1367 GGGCAGGAGGAGAAAGATGTGGGGCAGGAGGATGATGCCAGAGAGGAGGAG 1421

## RESULT 5

US-10-953-349-6314  
; Sequence 6314, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERBY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953.349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6314  
; LENGTH: 1147  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-10-953-349-6314

Query Match	1.4%;	Score 38.8;	DB 6;	Length 1147;
Best Local Similarity	47.2%;	Pred. No. 0.41;		
Matches 152;	Conservative 0;	Mismatches 167;	Indels 3;	Gaps 1;
Qy	674	AAGAGGAGGAGCGCTGATATATAAGCTAAAGGAGCGCTGCATCAGCTCGTTATGAAAC	733	
Db	549			
Qy	734	AAGAAGGAGTAAAAACATAGCCATCGATGCTTTTAAATTATGTGGGTCCGAGCTGATTGGCCAAA	793	
Db	609			
Qy	794	CGAGGCTCTGAGGACTGACAAAACAGAACCCAGGAATGAAGATGAGATCTACAAAATT	853	
Db	669			
Qy	854	CTGTTGATGATTATGAGTACCGCCAGACGACATCTGTATGGAGAACCGGAGCTGAAG	913	
Db	729			
Qy	914	AAGGTCCTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTGTCTCTCTCAGAGAAGAAG	973	
Db	786			
Qy	974	CCCAGGAAAGACGAGGAGCG	995	
Db	846	CACTATGATTAATGATGATGATG	867	

## RESULT 6

```

US-11-217-529-80932
; Sequence 80932, Application US/11217529
; Publication No. US200600959612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3

```

```

; SEQ ID NO 80932
;
; LENGTH: 3207
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80932

```

### Query Match

Query Match	1.48;	Score 38.8;	DB 7;	Length 3207;
Best Local Similarity	55.2%;	Pred. No. 0.76;		
Matches 73;	Conservative 0;	Mismatches 57;	Indels 0;	Gaps 0
Qy	839	GAGATGTACAAAAATCTGTTGAATGATTATGAGTACCGCCAGAACGAGATCCTGATGGAG	898	
Db	839	GTGNAGGACAGCATTTTTCAGAGGAGTAGAATGGAATTAGAAGACGACATTGATGGG	898	
Qy	899	AACCGCGAGCTGAAGAAAGTGCTCTCCAGCAGATGAAGAGGAGAGATGCATCTCTCTCTGTCT	958	
Db	899	AAAGCGATGCTGAAAAGAGATGAAAGTCAGGATGAAGAGGGAACAGACATTCGGTAGATT	958	
Qy	959	CCTCAGAA	968	
Db	959	TTTCAAAAA	968	

## RESULT 7

```

US-11-270-287-1/c
; Sequence 1, Application US/11270287
; Publication No. US20060099627A1
; GENERAL INFORMATION:
; APPLICANT: Kara, Anna K.
; APPLICANT: Ting, Robert C.
; APPLICANT: Tham, Jill M.
; APPLICANT: Nelson, James S.
; APPLICANT: Tan, Theresa M.
; TITLE OF INVENTION: Diagnosis of Parasites
; FILE REFERENCE: 64-99
; CURRENT APPLICATION NUMBER: US/11/270,287
; CURRENT FILING DATE: 2005-11-09
; PRIORITY APPLICATION NUMBER: US/09/369,992
; PRIORITY FILING DATE: 1999-08-06
; PRIORITY APPLICATION NUMBER: PCT/IB98/00212
; PRIORITY FILING DATE: 1998-02-05
; PRIORITY APPLICATION NUMBER: AU P09481/97
; PRIORITY FILING DATE: 1997-09-26
; PRIORITY APPLICATION NUMBER: AU P09329/97
; PRIORITY FILING DATE: 1997-04-21
; PRIORITY APPLICATION NUMBER: AU P04953/97
; PRIORITY FILING DATE: 1997-02-06
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5849
; TYPE: DNA
; ORGANISM: Plasmodium berghei
US-11-270-287-1

```

	Query Match	1.4%	Score 38.8	DB 7	Length 5849		
	Best Local Similarity	58.8%	Pred. No. 1.1				
	Matches	67	Conservative	0	Mismatches	47	
					Indels	0	
					Gaps	0	
Qy	2572	TGAGTTCAGATAC	TACTAGAACT	TACTACGTTAC	ATTTTGGAGATTTT	TAAAGCATTGTATT	2631
Db	726	TAATATAAAAT	TATTATGATAA	TAAATAAAAT	TAAAAAAT	TTTTTAAATATTTATAAT	667
Qy	2632	TTATTTTATATAT	GTGCAATGTT	ATATTTCT	TAAGAGAAAT	TATTCATTTATGGAGT	2685
Db	666	TTAAATTTAAATAT	CAATATATAAT	ATAAAAT	TTATTTATTTATTAAT	TAAGAAAT	613

## RESULT 8

US-11-257-851A-63/c  
; Sequence 63, Application US/11257851A  
; Publication No. US20060105951A1  
; GENERAL INFORMATION:

APPLICANT: Cunningham, Mark  
APPLICANT: Stojanovic-Susulic, Vedrana  
APPLICANT: O'Neill, Karyn  
APPLICANT: Huang, Chichi  
APPLICANT: Luo, Jeffrey  
TITLE OF INVENTION: Melanocortin Receptor Binding Mimeticibodies, Compositions, Methods  
TITLE OF INVENTION: Uses  
FILE REFERENCE: CENS080 USA NP  
CURRENT APPLICATION NUMBER: US/11/257,851A  
PRIORITY FILING DATE: 2005-10-25  
PRIORITY FILING DATE: 60/621,960  
PRIORITY FILING DATE: 2004-10-25  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 63  
LENGTH: 11978  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: In Vitro synthesized DNA with expression vector functions that  
FEATURE:  
OTHER INFORMATION: encodes an alpha-MSH mimetibody  
US-11-257-851A-63

Query Match 1.4%; Score 38.4; DB 7; Length 11978;  
Best Local Similarity 43.8%; Pred. No. 2.2; Mismatches 336; Indels 6; Gaps 2;  
Matches 266; Conservative 0; Mismatches 336; Indels 6; Gaps 2;  
QY 388 GAACGAGCTGCTCGTCTTACGCGGAAGAACCTGCTGCGCCAGGAGAGCGCTGGAGACACA 447  
DB 7070 GGAGCAGGAGCAGGAGCGGAGGGGCGCAGGACGAGGGGCGCAGGAGGAGGGGCA 7011  
QY 448 GAACTTGAAGCTGGGAGCTGACATGAGACCACTCGAGAGCTGCTACGCCAACTTAAGGA 507  
DB 7010 GGAGCAGGAGGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGG 6951  
QY 508 GCAGTTGAAACGCTCCAGGCGGAGAGTATCGGCTTCAAGAGAGACAGCAGCTGCA 567  
DB 6950 GCAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGGCA 6894  
QY 568 GTCAAGAAACAGGAGTTTGCATCAGCTCCTGAAGAATGAGAAAGATGAGGTACAAAAATT 627  
DB 6893 GGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGG 6834  
QY 628 ACAAATATCATAGCCAGCCGGCTACTCATGTAATAATCATGATGTAAGAGGAAGGAGCG 687  
DB 6833 GCAGGAGCAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAG-- 6776  
QY 688 TGAATATAAAGCTAAGGAGCGCTGCATCAGCTCGTTATGAACAAGAGGATAAAAA 747  
DB 6775 -GAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCA 6717  
QY 748 CATAGCCATGATGTTTTAAATATTGTTGGTTCGAGCTGATGGCAACAGAGGCTCATGGAG 807  
DB 6716 GGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCA 6657  
QY 808 GACTGACAAAACAGAGCCAGGAATGAAGATGAGATGTACAAAATTCCTTGAATGATTA 867  
DB 6656 GGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGCA 6597  
QY 868 TGAGTACCCAGAGCAGAGTCTTGATGGAGAACGCGAGCTGAAGAGGTCTCTCCAGCA 927  
DB 6596 GGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCA 6537  
QY 928 GATGAAGAAGGAGATGATCT 987  
DB 6536 GGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGCA 6477  
QY 988 AGAGACG 995  
DB 6476 GGAGCAGG 6469

RESULT 9  
US-10-953-349-23270  
; Sequence 23270, Application US/10953349  
; Publication No. US20060107345A1  
; GENERAL INFORMATION:  
; APPLICANT: ALEXANDROV, Nikolai et al.  
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES  
; TITLE OF INVENTION: ENCODED THERY  
; FILE REFERENCE: 2750-1579PUS2  
; CURRENT APPLICATION NUMBER: US/10/953,349  
; CURRENT FILING DATE: 2004-09-30  
; NUMBER OF SEQ ID NOS: 40252  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 23270  
; LENGTH: 1046  
; TYPE: DNA  
; ORGANISM: Glycine max  
US-10-953-349-23270

Query Match 1.4%; Score 38.2; DB 6; Length 1046;  
Best Local Similarity 54.7%; Pred. No. 0.57; Mismatches 0; Gaps 0;  
Matches 76; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 825 CCAGGAATGAAGATGAGATGTACAAAATTCCTTTGATGATTATGAGTACCGCCAGAACG 884  
DB 305 CCATGCTGAATGGGAGAGAACCAAGAAGATGGCGAACCCAGGAGAGCTTCCTAGGGCATA 364  
QY 885 AGATCCTGATGGAGAACCGCGAGCTGAAGAAGGTCTCTCCAGCAGATGAAGAAGAGATGA 944  
DB 365 TGATCGCAAGCAACAGGAGCAGCTGAAGAAGTCTGTGAAGGAAACAAAGAGGAGAGGA 424  
QY 945 TCTCTCTCTCTCTCTCTCTCA 963  
DB 425 TGACCGGTTCATGCATCA 443

RESULT 10  
US-11-301-554-1028  
; Sequence 1028, Application US/11301554  
; Publication No. US20060088527A1  
; GENERAL INFORMATION:  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Durham, Margarita  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McNabb, Andria  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
; FILE REFERENCE: 210121.478C21  
; CURRENT APPLICATION NUMBER: US/11/301,554  
; CURRENT FILING DATE: 2005-12-13  
; PRIORITY FILING DATE: US 10/283,017  
; PRIORITY FILING DATE: 2002-10-28  
; PRIORITY FILING DATE: US 10/113,872  
; PRIORITY FILING DATE: 2002-03-28  
; PRIORITY FILING DATE: US 10/017,754  
; PRIORITY FILING DATE: 2001-10-29  
; PRIORITY FILING DATE: US 09/902,941  
; PRIORITY FILING DATE: 2001-07-10  
; PRIORITY FILING DATE: US 09/849,626  
; PRIORITY FILING DATE: 2001-05-03  
; PRIORITY FILING DATE: US 09/736,457  
; PRIORITY FILING DATE: 2000-12-13  
; PRIORITY FILING DATE: US 09/702,705  
; PRIORITY FILING DATE: 2000-10-30

```

; PRIOR APPLICATION NUMBER: US 09/677,419
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/671,325
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/658,824
; PRIOR FILING DATE: 2000-09-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2157
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1028
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-301-554-1028

Query Match      1.4%;      Score 38;  DB 7;  Length 438;
Best Local Similarity 49.5%;  Pred. No. 0.38;
Matches 98;  Conservative 0;  Mismatches 100;  Indels 0;  Gaps 0;

Qy      591  AGCTCTCGAAGAATGAGAAAGATGAGGTACAAAAATTACAAAATATCATAGCCAGCCGGG 650
Db      211  AGGACTTAAAGGAGAAGAAGAAAGTTGTGGAAGAGGCCAGAAAATGGAAGAGACGCCCTG 270

Qy      651  CTACTCAGTATAATCATGATGTGAAGAGGAAGGCGTGAATATAATAAGCTAAAGGAGC 710
Db      271  CTAACGGGAATGCTAAATGAGGAAAATGGGGAGCAGGAGCGCTGCACAATGAGGTAGACGAAG 330

Qy      711  GCCTGTCATCAGCTCGTTATGAACAAGAGGATAAAAACATAGCCATGGATGCTTTTAAATT 770
Db      331  AAGAGGAGAAAGGTTGGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 390

Qy      771  ATGTGGCTCGAGCTGATG 788
Db      391  ATGGAGATGAAGATGAGG 408

```

```

RESULT 11
US-11-217-529-80879
; Sequence 80879, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 80879
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-80879

```

	Query Match	1.4%	Score 37.4	DB 7	Length 1434	
	Best Local Similarity	52.2%	Pred. No. 1.1			
	Matches	83	Conservative	0	Mismatches	76
					Indels	0
					Gaps	0
Qy	2116	GCCTCGGCTCTGTGACTGCAGGAATGCATGGCGCTTTGGATGGAACAGAAAGCGCTCGAA				2175
Db	652	GCCACACAGTGCCTTGAATGCAAGAAATGGTTCACTTTTAAAGAAAAAGTATCGCAGAA				711
Qy	2176	TGATTGCTCGCCAGGTACCGAAGAGACACTTTTAGGGACTGGTCTCTGTAAACATTAA				2235
Db	712	TCCTTTCATTCTCAACCAAGAAGACATTCGTTCTTTAGGGAATACATCAAGTACCAAGCA				771

```

Qy      2236 ATATTCTGCCCAAGTGTGGTTGTCATTTGGAAGTGTAGCC 2274
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      772 AAAAAAGCCAGRAGCGGTTCAGGAGAAATCTTGCC 810

RESULT 12
US-10-953-349-36772
; Sequence 36772, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THERAPY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36772
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
US-10-953-349-36772

Query Match           1.4%; Score 37.2; DB 6; Length 898;
Best Local Similarity 61.2%; Pred. No. 0.97; Indels 0; Gaps 0;
Matches 60; Conservative 0; Mismatches 38;

Qy      877 CCAGAAGCAGATCCTGTATGTAGAACCGGAGCTGTAAGAAGGTCCTCCAGCAGATGAAGAA 936
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      357 CCGGCTGCCGATGATGTGTCACGGCCGCAGAACGAGCACTTGTGCGGNACAGAC 416

Qy      937 GGAGATGATCTCTCTCTGTCTCTCTCAGAAAGAAGC 974
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      417 ACTGAACAACGCTCGTCCGGCTCTCAGAAAGAGCGAGC 454

```

```

RESULT 13
US-11-217-529-5994
; Sequence 5994, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIRO
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5994
; LENGTH: 3495
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5994

```

	Query Match	1.4%;	Score 36.6;	DB 7;	Length 3495;
	Best Local Similarity	46.6%;	Pred. No. 3.3;		
	Matches 117;	Conservative 0;	Mismatches 134;	Indels 0;	Gaps 0;
Qy	543	TTCAAGAGAGAGACAGCGCAGCTGCGACTGCAAGAACAGGAGTTTGTCATCAGCTCCTCAAGA	602		
Db	3089	TTGAGAAAGAAAGAAAGAGGAGAGCGAAGATAAGAAAGAAAGAAAGAGGAGAAGA	3148		
Qy	603	ATGAGAAAGATGAGGTACAAAAATTTACAAAATATCATAGCCAGCCGGGTACTCTAGTATA	652		
Db	3149	AGAAAGAAAAGAAAGGAAGAGGAGGAGGAAGAAAGAAAGAGGAAAGGAAAAGAGAAAGAAA	3208		





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:14:14 ; Search time 42 Seconds  
(without alignments)  
1408.888 Million cell updates/sec

Title: US-10-644-084-2  
Perfect score: 3165  
Sequence: 1 MGDWMTVPDLCTENKNLS.....CYSGSSAFPSAHGRDDLLP 615  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2786.5	88.0	614	2	T26467
2	238.5	7.5	1738	2	T14867
3	237	7.5	2331	2	T25410
4	234.5	7.4	1679	2	S48385
5	224	7.1	2139	2	T18296
6	223.5	7.1	3459	1	A56539
7	222.5	7.0	3225	2	I52300
8	220.5	7.0	1780	2	T17272
9	220.5	7.0	1875	2	S38173
10	220	7.0	2442	2	T08621
11	218.5	6.9	1128	2	G82666
12	218.5	6.9	1354	2	S74244
13	218.5	6.9	1538	2	T29095
14	217.5	6.9	746	2	T47237
15	217.5	6.9	1909	2	A45592
16	215.5	6.8	1388	2	S74245
17	215.5	6.8	2649	2	A40937
18	214	6.8	980	2	E71606
19	214	6.8	1085	2	F96712
20	213	6.7	944	2	S26710
21	213	6.7	1955	2	T30934
22	212.5	6.7	2663	1	S28261
23	212	6.7	1790	2	S67593
24	212	6.7	1979	1	S03166
25	211.5	6.7	1181	2	T01799
26	211.5	6.7	1388	2	S70633
27	211	6.7	2253	2	T30336
28	210.5	6.7	1354	2	S69211
29	210	6.6	911	2	SS1441

ALIGNMENTS

RESULT 1

T42649  
hypothetical protein DKFZp434C0515.1 - human  
C:Species: Homo sapiens (man)  
C>Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T42649  
R:Blöcker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, November 1999  
A:Reference number: Z22230  
A:Accession: T42649  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-614 <AAA>  
A:Cross-references: UNIPROT:O9UIX0; UNIPARC:UP1000006DABF; EMBL:AL133046  
A:Experimental source: adult testis; clone DKFZp434C0515  
C:Genetics:  
A>Note: DKFZp434C0515.1

Query Match		88.0%	Score 2786.5;	DB 2;	Length 614;
Best Local Similarity		87.5%	Pred. No 3e-127;		
Matches 538;		Conservative 35;	Mismatches 41;	Indels 1;	Gaps 1;
QY	1	MGDMWMTVPDLCTENKNLSQVTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60		
DB	1	MGDMWMTVPDLCTENKNLSQVTSKMPSSLYSQVLCSSVPLSKNVHGVFGVCTGE	60		
QY	61	NIEQSIYLDQELTTFGPSPSYEESKSKAKRELNIIVAVLNCWELLVLRKNLLAQESV	120		
DB	61	NIEQSIYLDQELTTFGPSPSYEESKSKAKRELNIIVAVLNCWELLVLRKNLLAQENV	120		
QY	121	ETQNLKLGSDMDHLQSCYAKLKEQLTSRRREMIGLQERDRLQCKNRSLLHLLKNEKDEV	180		
DB	121	ETQNLKLGSDMDHLQSCYAKLKEQLTSRRREMIGLQERDRLQCKNRLHLLKNEKDEV	180		
QY	181	QKLQNIIRATQYNHDKRKEREYNKLERLHQLVMNKDKNIAMDVLNYYVGRADGKRG	240		
DB	181	QKLQNIIRATQYNHDKRKEREYNKLERLHQLVMNKDKKIAMDILNYYVGRADGKRG	240		
QY	241	SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVLOOMKEMISLLSPQKKPR	300		
DB	241	SWRTDKTEARNEDEMYKILLNDYEQKQILMENAEKLVLOOMKEMISLLSPQKKPR	300		
QY	301	ERAEDGTGTVAISDIEDDSGELSRDSVWGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ	360		
DB	301	ERVDDSTGTV-ISDVEEDAGELSRDSVWGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ	359		
QY	361	ASKVHSEGLNEEDVTSRDHEQETEKLEIERCKEMIKAQQLLQQLATTCDDDTTSL	420		
DB	360	VSKVHLEGFNEDVTSRDHEQETEKLEIQCKEMIKTQQQLLQQLATAYDDTTSL	419		
QY	421	LRDCVLLBEKRLKEEWTLFKQEKQNFRRRRSFTAAAIRGLERKATFEERASWVKQOF	480		



Db 2292 SRISTA 2297

RESULT 4  
S48385  
hypothetical protein YIL149c - Yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 09-Jul-2004  
C:Accession: S48385  
R:Churcher, C.  
submitted to the EMBL Data Library, September 1994  
A:Reference number: S48310  
A:Accession: S48385  
A:Molecule type: DNA  
A:Residues: 1-1679 <CHU>  
A:Cross-references: UNIPROT:P40457; UNIPARC:UPI000013B453; GB:Z47047; EMBL:Z38059; NID:9  
A:Gene: SGD:MLP2; MIPS:YIL149C  
A:Cross-references: SGD:S0001411  
A:Map position: 9L

Query Match 7.4%; Score 234.5; DB 2; Length 1679;  
Best Local Similarity 19.6%; Pred. No. 0.00087;  
Matches 145; Conservative 124; Mismatches 230; Indels 241; Gaps 28;

QY 15 ENKLSQYT-----SETKSPSSLYSOQVLCSSVPLSKNVHGVFGVCTGE--- 60  
Db 1010 ENKJSELVIRLEKDAADCAELTKTKSSLYSAQDL-----LDKHKRWMEEKADYEREL 1064

QY 61 --NIEQS-----ISVLDQELTFG-----PPSLYBESKSEAKRELNIVA 98  
Db 1065 ISNIEOTESLRVENSVLIEKVDTAANGDKHLKVLVSFLSNLRHNSLET----- 1117

QY 99 VLNCMNELLVLRKNNLLAQESV-ETQNLKLGSDMDHLOSCYAKLKEQLETSRREMIG--- 154  
Db 1118 LTTCKRELAFVAKQKDSLEINDLQRTQTLSEKEY--QCSAVIIDEFKDITKEVTQVNI 1175

QY 155 -----LQERDQLOCKRSLHLLKNEKDEQVKLONIIASRATQY---NHDVKKERE 204  
Db 1176 LKENNAILOKSLKVNTEKREIYKQLNDQREISRLQDLIQTKQVSNKILVYSE 1235

QY 205 YNKLKERLHOLVMNKKKNIAAMDVLNVYGRADKRGSRWTDKTEARNEDEMYKILLNDYE 264  
Db 1236 MEQCKORYODLSQQKD-----AQKQIEKTNE-----ISDLK 1269

QY 265 YRQKQILMENAEKLVLOQMKKEMISLLSPQKKPRERAEDGTGTVAISDIEDDSGELSR 324  
Db 1270 GKLSAENANADLENKFNLRKQAHEKLDASKQ----- 1303

QY 325 DSVWGLSCDTVREQLTNSRQWRILKSHVEKLDNOA-----SKVHSEGLNEED 373  
Db 1304 -----QAALTNELNE---LKAIKDKLEQDLHFENAKVIDLDLTKLKAHELQSD 1348

QY 374 VISRODHEOET-----EKLELETERCKE-MIKAAQOOL 404  
Db 1349 V--SRDHEKDTVRLMEETESLKKELQIFKTANSSSDAFEKLVNNEKEDRIIDERTKE 1406

QY 405 LQOQLATTCDDDTSLRLDCYLLEKERLKEEW-----TL--FKEQKKNFERRRSFT 455  
Db 1407 FEKKLQETLNKSTSS--BAEYSKDLETLKKEWLKEYEDETLLRIKEAENLKKRILRPS 1463

QY 456 EAAIR--LGLERKAFEEERASVWQO----- 479  
Db 1464 ERIQKITSKRKEELEEFKRLKENAGSLTFLDNKSGEDAEELWNSSPKSGNSERPSPA 1523

QY 480 ---FLNMTNFDHQS--ENVKLFSAFGS-----SDPDNLI VHSPRQKKLHVANGVPAC 530  
Db 1524 VAGFINQKMLKQEQKLVKNVNDVFNDSQSMVTNKENNIVDSAAAGNKAIPTFSGKPPF 1583

QY 531 TSKLTKSLPASPSTDFRTHSCVSHSSISVINITPEE-----SKPSEVARESTDQ 582  
Db 1584 SSN-TSSLQS--FQNPFTASQSNINTNAPLRLTINTLOPEVAVKAAINFSNVTDLTNNSTDG 1640

QY 583 KWSVQRSPSSREGCYSGCSS 602  
Db 1641 AKITEIGSTSKRPIESGTSS 1660

RESULT 5  
T18296  
myosin heavy chain - Entamoeba histolytica  
C:Species: Entamoeba histolytica  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18296  
R:Guillen, N.  
submitted to the EMBL Data Library, February 1997  
A:Reference number: Z18865  
A:Accession: T18296  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2139 <GUI>  
A:Cross-references: UNIPROT:Q07569; UNIPARC:UPI00000802D3; EMBL:L03534; NID:91850912; P  
C:Genetics:  
A:Gene: mhca  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
F:91-780/Domain: myosin motor domain homology <MMO>

Query Match 7.1%; Score 224; DB 2; Length 2139;  
Best Local Similarity 20.5%; Pred. No. 0.0037;  
Matches 158; Conservative 110; Mismatches 240; Indels 262; Gaps 32;

QY 15 ENKLSQYTSETKMPSSLYSOQVLCSSVPLSKNVHGVFGVF-----CTGE---NIE 63  
Db 1100 EBDTSLSONLEKLTNTKTK-----TKADLEKKISGLQDYEDLEDDKNNKIEGDLRNAQ 1154

QY 64 QSIYLDQELTTFGFPPLY-----BESKSEAKRELNIVAVLNCMELLVLRKNNLLAOE 118  
Db 1155 RKIKELDDIITGADVSVYLOKQKEYESQIAKMQEKEAIGND-----VKNKEKTIKEK 1209

QY 119 SVETQNLKLGSDMDHLOSCYA-KLKEQLETSRREMIGLOERDQLOCKRSLH-----QLL 173  
Db 1210 ELEIQSLOEKLDETEVEKEDAEKKKEIE---KEMKALQEEKENVENSSKTEKDKKLE 1266

QY 174 KNEKDEVOKLONIIAS-----RATQVNDV-----KPK-----EREYN 206  
Db 1267 DNLKQTKQLDDMTADNEKLIKAKADLEAQLNEVQDNHEKAVADAEKLNKKAQSKELN 1326

QY 207 KLKERLHQLV-----MNNKDKNIAMDVLNVYGR----- 234  
Db 1327 SLKAELEALTAKSVESKNKDSENEKAALSEEIDQANEKKNIQADLRKATADLQEAENE 1386

QY 235 -----ADKRGSRWTDKTEARNEDEMYKILLNDYE----- 264  
Db 1387 KKAERVEAQRDKLVADNKKMTKTLEIKARDEENTYKV--ENYEKVLKRKEADLEEAENL 1444

QY 265 -----YRQKQILMENAEKLV-----LQOMKK----- 286  
Db 1445 DIEKQDNWKEQVKLEGELETKDKLNAATAEKDSIPTAKKQSDADLEELNKTVEEHD 1504

QY 287 EMISLLSPQ-----KKKPRERAED-----GTGT--- 309  
Db 1505 EVVAKLNTQITKLTRDNQSAEBELNELRSKADKKKISELEEQVNELESRPVGTGNAD 1564

QY 310 --VAISD--IEDDSGELSRDSVWGLSCDTVRQL---TNSIRKQWRILKSHVEKLDN--- 359  
Db 1565 NEIKIRDAQIADLNKALEMKGVQNNLOQATNKLKAKNDLTSKIEITENEMKKLENAKK 1624

QY 360 --QASKVHSEGLNEEDVISRODHEOETKLELEIERCKEMIKAAQOOLLQOQLATTCDDDT 417  
Db 1625 RLEQDKDADKAVSQTIKRKGLEEEVKKLTTEIQALFKQINAPSSVAQE----- 1674

QY 418 TSLLRDCYLLBEKERLKEEWTLFKQKKNFERRRSFTAAAIRGLERKAFEEERASVWK 477  
Db 1675 -----BEKQRLSDIAELKEQ---LEQERTTAANAB---AERKKIQAE-LDEVK 1716

QY 478 QOFLNMTN-----FDHQNSENKVLPSAFSGSSDPDNLIVHSPRQKKLHVANGVPACTSK 533

Db 1717 FNLEDVNTQREKLVAKNSN-----DABIDSL-----KEEKALEDETEKITDDNNK 1763

QY 534 LTKSLPASPSTSDPFRQTHSCVSEHSISVLNITPESKPSVARESTQOK 583

Db 1764 LSEEI-----DSLDRKYNALLDSKSDVSMKEKFQDELKVTKDALE-TEKK 1808

RESULT 6

A56539

giantin - human

N:Alternate names: macrogolgin

C:Species: Homo sapiens (man)

C>Date: 19-Oct-1995 #sequence\_revision 26-Jan-1996 #text\_change 09-Jul-2004

C:Accession: A56539; S57536

R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M. Mol. Cell. Biol. 14, 2564-2576, 1994

A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein

A:Reference number: A56539; MUID:94187728; PMID:7511208

A:Accession: A56539

A:Molecule type: mRNA

A:Residues: 1-3259 <SEE>

A:Cross-references: UNIPROT:Q14789; UNIPARC:UPI000012B42C; EMBL:X75304; NID:9405714; PID F:3238-3254/Domain: transmembrane #status predicted <TMN>

C:Genetics:

A:Gene: GDB:COLGB1; GCP: GCP371

A:Cross-references: GDB:454958

A:Map position: 3q13.31-3q13.31

C:Superfamily: giantin

C:Keywords: coiled coil; Golgi apparatus; transmembrane protein

F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 7.1%; Score 223.5; DB 1; Length 3259;

Best Local Similarity 19.9%; Pred. No. 0.0065;

Matches 147; Conservative 136; Mismatches 218; Indels 237; Gaps 34;

QY 15 ENKNLSQYTSPTKMS-----PSSLYSQVLCSSVP-LSKNVHGVFGVCTGENIEQ 64

Db 1225 ENENIGDQLRQLQVRESIDGKLPT--DQESCSSTPGLEELPFKATEQHHTQPVLES 1282

QY 65 SI-----SYLDELTTTFPPSLYE---ESKSEKARELNIVAVLNCMELLVLQKKNLLA 116

Db 1283 NLCPWPMSHSDASALOGGTSVAQIKAQKIEAEKVELEKVSSTTSELTKKSEVFOL 1342

QY 117 QESVETQNLKGS--DMDHLOSCYAK-LKEQLETSSRRMIGLQERDROLQCKNRSLSHQLL 173

Db 1343 QEQINKQGLEIESLKTVSHEAEVHAESLQOKLESSQLIAGL-EHLRELQPK----- 1393

QY 174 KNEDEVOKLQNIIASRATQYNHDKREREYNKLERHOLVMNKKDKNIAMDVLNVYG 233

Db 1394 ---LDELQKL-----ISKKEEDVSYLSGQL-----SEKEAALTKI----- 1425

QY 234 RADKRGSRWTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK----- 278

Db 1426 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLOVELCEMKQKPEEIGE 1473

QY 279 --KVLQMKKEMISLLSPQK-----KKPRERAEDGTGTV-----AISDIED--DSGELS 323

Db 1474 ESRAKQIQRLQQAALISRKALKENKSLQELSLSLARGTIERLTYSKLDVSVQSAQNK 1533

QY 324 RDSVWG-----LSCDTVR-----BOL 339

Db 1534 KDTVLGRLLAQEERDKLITEMDRSLLENQSLSSCSLSKLALEGLTEDKEKLVKEIESL 1593

QY 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVLRS 377

Db 1594 KSSKIAESTEWQEKHKEQYELLLOSVENVSNAEARIQHVVEAVRQEKQELYGKIRST 1653

QY 378 QDHEQETEK---LELETERCKEMIK-----AQOQLLOQ-----LATTC 413

Db 1654 EANKKETEKQLEAQEQEEMENKMKRFAKSKQKQKILEEENDRLRAEVHPAGDTAKEC 1713

QY 414 DDDTTSLLRDCVLLKEERLKEEWTLPKEQKKNFERRRSFTEAARIGLERKAFEBERA 473

Db 1714 ME--TLSSNASMKEELERVRMEYETLSKFKQSLMSEKDSLSSEVQDL-----KHQIEDNV 1767

QY 474 SWVKQQLNMTNFDHQNSNVK---LFSAFSGSSDPDNLIVHSRPRQKKLHSHVANGVPAC 530

Db 1768 S--KOANLEATE-KHDNQTNVTBEGTQIPQETBEQDSLSMSTR-----PTC 1811

QY 531 TSKTUKSLPASPSTS-DPROTHSCVSEHSSIS-----VLNITPESKPSSEVAR 577

Db 1812 SESVPSAKSANPAVKQF-----SSHDEINNVLIQDQLKERIAGLEEBKQKKNKFSQ 1864

QY 578 ESTQKQSVQSRPSREG 595

Db 1865 TLENKNTLLSQISTKDG 1882

RESULT 7

152300

giantin - human

N:Alternate names: gcp372

C:Species: Homo sapiens (man)

C>Date: 01-Nov-1996 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004

C:Accession: 152300

R:Sonda, M.; Misumi, Y.; Fujiwara, T.; Nishioaka, M.; Ikehara, Y. Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994

A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in

A:Reference number: 152300; MUID:95100974; PMID:7802676

A:Accession: 152300

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3225 <RES>

A:Cross-references: UNIPROT:Q14789; UNIPARC:UPI000016A987; GB:D25542; NID:9662389; PIDN C:Superfamily: giantin

Query Match 7.0%; Score 222.5; DB 2; Length 3225;

Best Local Similarity 18.9%; Pred. No. 0.0071;

Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;

QY 15 ENKNLSQYTSPTKMS-----PSSLYSQVLCSSVP-LSKNVHGVFGVCTGENIEQ 64

Db 1191 ENENIGDQLRQLQVRESIDGKLPT--DQESCSSTPGLEELPFKATEQHHTQPVLES 1248

QY 65 SI-----SYLDELTTTFPPSLYE---ESKSEKARELNIVAVLNCMELLVLQKKNLLA 116

Db 1249 NLCPWPMSHSDASALOGGTSVAQIKAQKIEAEKVELEKVSSTTSELTKKSEVFOL 1308

QY 117 QESVETQNLKGS--DMDHLOSCYAK-LKEQLETSSRRMIGLQERDROLQCKNRSLSHQLL 173

Db 1309 QEQINKQGLEIESLKTVSHEAEVHAESLQOKLESSQLIAGL-EHLRELQPK----- 1359

QY 174 KNEDEVOKLQNIIASRATQYNHDKREREYNKLERHOLVMNKKDKNIAMDVLNVYG 233

Db 1360 ---LDELQKL-----ISKKEEDVSYLSGQL-----SEKEAALTKI----- 1391

QY 234 RADKRGSRWTDKTEARNEDEMYKIL-----LNDYEVROKQILMENAELK----- 278

Db 1392 -----QTEIIEQEDLIKALHTQLEMQAKEHDERIKQLOVELCEMKQKPEEIGE 1439

QY 279 --KVLQMKKEMISLLSPQK-----KKPRERAEDGTGTV-----AISDIED--DSGELS 323

Db 1440 ESRAKQIQRLQQAALISRKALKENKSLQELSLSLARGTIERLTYSKLDVSVQSAQNK 1499

QY 324 RDSVWG-----LSCDTVR-----BOL 339

Db 1500 KDTVLGRLLAQEERDKLITEMDRSLLENQSLSSCSLSKLALEGLTEDKEKLVKEIESL 1559

QY 340 TNS-----IRKQWRILKSHVEKLDNQASKVHS--EGLNEE-----DVLRS 377

Db 1560 KSSKIAESTEWQEKHKEQYELLLOSVENVSNAEARIQHVVEAVRQEKQELYGKIRST 1619

QY 378 QDHEQETEK---LELETERCKEMIK-----AQOQLLOQ-----LATTC 413

Db 1620 EANKKETEKQLEAQEQEEMENKMKRFAKSKQKQKILEEENDRLRAEVHPAGDTAKEC 1679

```
QY 414 DDDTTSLRDCYLLERKLEKEWTLFKEQKNFERERSFTEAARLGLERKAFEBERA 473
Db 1680 ME--TLLSSVASWKEELERVVMYETLSKRFQSLMEKOSLSBEVDLKHQIEGNSVSKQA 1737
QY 474 SM-VKQOFLMNTFDRHNSENVLFSAFSGSSDPDNLIVHSRPRQKKLHLSVANGVPACTS 532
Db 1738 NLEATEKHNDQNTVTEGTQISF-----GTEBQDLSMSTR-----PTCSE 1779
QY 533 KLTKSLPASPSTS--DFRQTHSCVSEHSSIS-----VNLTPBESKPSVARBS 579
Db 1780 SVPSAKSANPAVSKDF-----SSHDEINNYLQIDQLKRIAGLEEBEKQKNKFSQTL 1832
QY 580 TDQKWSVQSRPSSREG 595
Db 1833 ENKKNLTLISQISTKOG 1848

RESULT 8
T17272
hypothetical protein DKFZp434B0435.1 - human
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17272; T46451; A36881
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: Z18723
A:Accession: T17272
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1780 <POU>
A:Cross-references: UNIPROT:Q9UFRS; UNIPARC:UPI0000037D09; EMBL:AL117496
A:Experimental source: adult testis; clone DKFZp434B0435
R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23028
A:Accession: T46451
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 575-1136, 'C', 1138-1608, 'K', 1610-1780 <AAA>
A:Cross-references: UNIPARC:UPI000006D421; EMBL:AL1137392
A:Experimental source: adult testis; clone DKFZp4341152
R:Westendorf, J.M.; Rao, P.N.; Gerace, L.
Proc. Natl. Acad. Sci. U.S.A. 91, 714-718, 1994
A>Title: Cloning of cDNAs for M-phase phosphoproteins recognized by the MPW2 monoclonal
A:Reference number: A36881; MUID:94119956; PMID:8290587
A:Accession: A36881
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1215-1261, 'E', 1263-1418, 'S', 1420-1608, 'K', 1610-1780 <WES>
A:Cross-references: UNIPARC:UPI000016AD74; GB:L16782; NID:G292328; PIDN:AAC37542.1; PID:
C:Genetics:
A>Note: DKFZp434B0435.1; DKFZp4341152.1
C:Keywords: phosphoprotein

Query Match 7.0%; Score 220.5; DB 2; Length 1780;
Best Local Similarity 20.6%; Pred. No. 0.0044;
Matches 156; Conservative 117; Mismatches 235; Indels 251; Gaps 36;

QY 16 NKNLSQYTSRQWSPSSLYS-----QQVLCSSVPLSKNVHGVFGVCTGENTEQSI 66
Db 996 SKQVKEYRIQEPNRENSSHSTAEATWECKEIVKAS---SKSHQI-----EELEQOI 1045
QY 67 SYLDQELFTTFGPPSYEESKSKAKRELNIIVAVLNCMNBELLVLQRNLLAQBS-----VET 122
Db 1046 EKLOAEVK--GKYDENNRKLEKHNQDDLLK---EKETLLQQLKEELQEKKNVLDVQI 1099
QY 123 QNLKLG-----SMDHLQSCY-AKLKEQ---LETSR-----REMIGLQERD 159
Db 1100 QHVVEGKALSELGTQVCTYKAKIKELETTLETQKVERSHSAKLEQDILEKESI-ILKLE 1158
QY 160 ROLQCKNRSLLHLLKNKXD-----
Db 1159 RNLKFEQHLQDQSVNTKDLNVKELKLBKEITQLTNLQDMKHLQLLQKBEETNRQETE 1218
```

```
QY 182 KL-QNIIASRAQYNN--HDVKKREBYNKLKERLHQLVMNKKDKNITAMDVLNVYGRADGK 238
Db 1219 KLKEELSASSARTQNLKADLQKEEDYADLKEKJTD--AKQIKQVQKEV-----1266
QY 239 RGSWRDTKTEARNEDEMYKILLNDVEYRQKQILMENAELKKVLQOMKKEMISLLSPQKKK 298
Db 1267 -----SVMRDEDKLRIKINELEKKKNQCSQELDMKQRTIQQLKEQLNNQKVEAIIQ 1318
QY 299 PRERA-----EDGTGTVAISDIEDDGSLSRDSVWGLSCDTVREQLTNSIRKOW 347
Db 1319 QYERACKDLNVKEKIIEDMRMTLE-----EQEQTQVEQDQVLEAKLEEV-ERLATELEK-W 1372
QY 348 RILKSHVEKLDNOASKVHSEGLNEEDVISR-----QDHEQETEKLELEIERCKEMIKAOQQ 403
Db 1373 KEKCNLDLETKNORSNKEHE--NNTDVLGKLTLNQLDELQESQ-KYNADR-KKWLEKWM 1428
QY 404 LLQOQLATTCDDDTTSLRDCYLLBEKERLKEEWTLFKEQKNFERERSF-----454
Db 1429 LITQ-----AKEAENIRN-----KEMKKYAEDRERFFKQONEMEIL 1464
QY 455 TEAAIRLGLERKAFEBER-----ASWVKOOFNLMTNFDHON-----SENVKLFSAF 500
Db 1465 TQALTEKSDLDQKWEERDQVAALEIQKALISSNVQKDNIEQLKRIISETSKIETQI 1524
QY 501 S-----GSSDPDNLIVHSRPRQKKLHLSVANGV-----527
Db 1525 MDIKPKRISSADPDKL--QTEPLSTSFSEISRNKIEDGSSVVLDSCEVSTENDQSTRFPKPE 1582
QY 528 -----FACTSKLTKSLPASSTSDFRQTHSCVSEHSSI--SVNLITTP 567
Db 1593 LEIQFTPLQPNKMAVKHPGCTTPVTVEIPKA-----RRKKGMEEDLVKCNKKNATP 1636
QY 568 EBSKE---SEVARESTDOKWSVQSRPSSREGCVSGCSSA 603
Db 1637 RTNLKFPISDDRNSSVKKEQKVAIRPSSKK-TYSLSRQA 1674

RESULT 9
S38173
myosin-like protein MLPI - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKR095w; protein YKR415
C:Species: Saccharomyces cerevisiae
C>Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 09-Jul-2004
C:Accession: S38173; S40647; S31207
R:Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38158
A:Accession: S38173
A:Molecule type: DNA
A:Residues: 1-1875 <BAL>
A:Cross-references: UNIPROT:Q02455; UNIPARC:UPI0000053164; EMBL:Z28320; NID:G486586; PI:
A:Experimental source: strain S288C
R:Bou, G.; Esteban, P.F.; Baladron, V.; Gonzalez, G.A.; Cantalejo, J.G.; Remacha, M.; J.
Yeast 9, 1349-1354, 1993
A>Title: The complete sequence of a 15 820 bp segment of Saccharomyces cerevisiae chrom
A:Reference number: S40644; MUID:94205265; PMID:8154186
A:Accession: S40647
A:Molecule type: DNA
A:Residues: 1-1875 <BOU>
A:Cross-references: UNIPARC:UPI0000053164; EMBL:X73541; NID:G450550; PIDN:CAAS1948.1; PI:
A:Experimental source: strain S288C
R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A>Title: A new yeast gene with a myosin-like heptad repeat structure.
A:Reference number: S31207; MUID:93247549; PMID:8483450
A:Accession: S31207
A:Molecule type: DNA
A:Residues: 1-300, 'A', 302-1875 <KOE>
A:Cross-references: UNIPARC:UPI0000168CE6; EMBL:L01592; NID:G171958; PIDN:AAA34783.1; PI:
C:Genetics:
A:Gene: SGD:MLPI
A:Cross-references: SGD:S0001803; MIPS:YKR095w
```

A;Map position: 11R

Query Match 7.0%; Score 220.5; DB 2; Length 1875;  
Best Local Similarity 20.2%; Pred. No. 0.0047; Mismatches 238; Indels 171; Gaps 27;  
Matches 134; Conservative 119; Mismatches 238; Indels 171; Gaps 27;  
QY 16 NKNLSQYTSKSPSSLYSQ-QVLCSS---VPLSKNVHGVGVFCT-----58  
DB 1107 NSRIEDLSQNKL-----LYDQIQIYTAADKEVNVNSTNGPLNNILTLRRERDILDTKVT 1162  
QY 59 -----GENTIEQISYLDQELTTFGFSPSYEESKSK-----BAKRELNIIVA-----VLNOM 103  
DB 1163 VAERDAKMLRQKISLMDVEL-----QDARTKLDNSRVEKENHSSIIQQHDDIMEKL 1213  
QY 104 NELLVLQRNLLAQESVETONLK---LGSMDHLQSCYAKLKEQLETSRPREMIGLOERDR 160  
DB 1214 NQNLRLRESNITLRNELENNNNKKELQSELQKLNQVAPISELTALKYSM---OEKEQ 1270  
QY 161 QIQCKNRSJHQLLKNEKDEVQKLNIIASRATQYNHDKREYNNKLERJHQLVMMNK 220  
DB 1271 ELKLAKEEVHRWKKRSQDILEKHEQLSSS-----DYEKLESEIENLKEELE---NKE 1319  
QY 221 DKNI-AMOVNLVYGRADGRGSRWRTDKTEARNEDEMYKILLNDYEVROKQIILMENAEKK 279  
DB 1320 RQGAEAEEKFNRLRQOQER--LKTSKLSQDSLTEQVNSLRDAKNVLENSLSEANARIEE 1377  
QY 280 VLQOKKEMISLLSPOKKPRERAEDGTGTVAISDIEDDSGELSRDSVWGLSCDVTVR-EQ 338  
DB 1378 -LONAKV-----AQGNQLEAIRKLQEDAERKASRELQAKLEESTTSYES 1420  
QY 339 LTNSTRKQWIRLKSHVEKLDNQASKVHSEGLNEEDVISQDHEQETEKLEIEIERCKEMI 398  
DB 1421 TINGLNEEITLKEELEKQIQOQLOQATSANEQNDLS---NIVESMKKSPEEDKIK-FI 1476  
QY 399 KQAOQLLOQOQATTCDDTTSLRLDCYLLEEKERLKEEWTL-FKQKQKNFERER-----451  
DB 1477 KEKTOEVNEK-----ILEAQERLNQPSNINWERIKKKWSEHEQEVSQ 1519  
QY 452 -----RSFTEAARLGLERKAFEBERASWVKQQLFNMTNPDHQNSENKLFPS 498  
DB 1520 KTRAEAEALKKIRLPTBEKINKIRKKEELEK-----EFEEKVEERIKSWE 1567  
QY 499 AFGSGSD---PDNLIVHSPROKKLHSAVANGVPACTSKTLKSLPASPSTS-----545  
DB 1568 Q-SGEIDVVLKQLEAKVQEKQKLENEYN-----KKLQEBELKDVPHSSHISDDERDKL 1620  
QY 546 -----DPRQTHSCVSEHS-----SISVLNITPBESEKSEVARESTDQKWSVQSRP 590  
DB 1621 RAEIESRLREEFNNELQAIKKKSFDEGKQOAMMKTTLLERKLAKMESQLSETKQSAESPP 1680  
QY 591 SS 592  
DB 1681 KS 1682

RESULT 10  
T08621  
centrosome associated protein CEP250 - human  
C;Species: Homo sapiens (man)  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: T08621  
R;MacK, G.J.; Rees, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.  
Arthritis Rheum. 41, 551-558, 1998  
A;Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera read  
A;Reference number: Z16462; MUID:98165428; PMID:9506584  
A;Accession: T08621  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-2442 <MAC>  
A;Cross-references: UNIPARC:UPI0000071A1C; EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1;  
A;Experimental source: cell line HeLa

Query Match 7.0%; Score 220; DB 2; Length 2442;

Best Local Similarity 21.6%; Pred. No. 0.0068;  
Matches 130; Conservative 103; Mismatches 181; Indels 188; Gaps 28;  
QY 60 ENIQSISYLDQELTTFGFSPSYEESKREKRELNIIVAVLNCNELLVQKKN-LLAQE 118  
DB 1490 EHLPMVAQVERQKLTQV-QREQIREPEKDRQTR--NVLE-----HOLLEKLEKQDMIESQ 1541  
QY 119 SVETONLK-----LGSMDM---HLQSCYAKLKEQLETSRREM-----IGLOERD 159  
DB 1542 RGQVODLKKQLVLTLECLALELEENHHKMECOOKLIKELEGORETORVALTHLTLEERS 1601  
QY 160 RLQOLCKNRLSHOL-----LKNKEDEVQKLNIIASRATQYNHDKRE 202  
DB 1602 QELQAQSQIHDLSEHSVTLARELQERDQEVKSQREQIEELQR-----QKEHLTQDLERRD 1657  
QY 203 REYNKLERLHQLVNNKKDKNIAMDVLNVYGRADGRGSRWRTDKTEARNEDEMYKILLND 262  
DB 1658 QELMLQKERIQ-----VLED 1672  
QY 263 YEYRQKQIILMENAE-LKKVLQOMKKEMISLLSPOKKPRERAEDGTGTVAISDIEDDSGE 321  
DB 1673 QRTQTKLLEEDLEQIKLSLRGRG-----LTTQRLMQERAEEGKGP---SKAQRGSLE 1725  
QY 322 ----LSRDSVWGLSC-----DTVREQLTNSIRK--QWRILKSHVEK-----LDN 359  
DB 1726 HMKLILRDKEREVECCQSHIHELQELKDQLEQQLGLHRKVGETSLLLSQREQEIVWLQQ 1785  
QY 360 QASKVHSEG-LNEEDVTISROD-----HEQETEKLELE-----IERCKEMIKAAQQ 403  
DB 1786 QLQEAREQELGELQSLOSQDEARALAQRODELEALQEQEQOQAQGBERVKEKADALQG 1845  
QY 404 LLOQOQATTCDDTTSLRLDCYLLEKEE---RLKEEWTL-----PKE 442  
DB 1846 ALQQAHTLKE-----RHGELQDHKEQARLEELAEVGRVQALBEVLGDLRAESRE 1898  
QY 443 QKKNFERERRSFTAAIRLGLERKAFEBERASWVKQQLFNMTNPDHQNSENKLFPSASG 502  
DB 1899 QEKALLALOQOQCAEQAEQHEVETRALQD---SWLQAQAVL-----KERDQLEALRAESQ 1950  
QY 503 SSDPDNLIVHSPRPQ-----KKLHSAVANGVPAC---TSKLTKSLPASPST-----SDFRQT 550  
DB 1951 SSRHQEAARAFARALQALQAKAHAAQKGKQHLLEQAELSRSUEASTATLQASLDACQA 2010  
QY 551 HS 552  
DB 2011 HS 2012

RESULT 11  
G86266  
hypothetical protein F3P19.25 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: G86266  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86266  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1128 <STO>  
A;Cross-references: UNIPROT:Q9SAF6; UNIPARC:UPI00000A7BC7; GB:AE005172; NID:g4850405; PI  
C;Genetics:  
A;Map position: 1

Query Match 6.9%; Score 218.5; DB 2; Length 1128;  
Best Local Similarity 23.4%; Pred. No. 0.0032;  
Matches 135; Conservative 108; Mismatches 196; Indels 139; Gaps 30;

QY 11 VLCTENKLSQYTSSETKMPSSLYSQVLCSVPVLSKNVHGVFGVCTGENI-----E 63  
DB 105 LLLMENKEL--VSKHEQLNOAFQAEIL--KREQSSHLYALTTVQREENLRKALGLEK 160  
QY 64 QSIYSYLDQBLTTGFPSPVLESK-----SKEAKR-ELN-IVAVLNCWN---ELLVLQRK 114  
DB 161 QCVOELEKALR-----EIOENSKIRLSSEAKLVEANALVASVNGRSSOVENKIYSAESK 215  
QY 115 LAQESVETQNLK-----GSDMDHLQSCYAKLKEQLE-ISRREMIGLQERDRLOCLK 165  
DB 216 LAEATRKSELKRLKEVETRESVLOQERLSFTKRESYEGTFQKQREYLNEWEKLG 275  
QY 166 NRSLLHLLKN-----EK-DEVQKQLNIIASRATQYNHNV-----KREKREYNKLERLHQ 214  
DB 276 EESITEQKLNQREKVEIEKKLKEKELEEMNRKVDLSKSKETE-EDITKLEEE 334  
QY 215 LVNKKDKNIAMDVLNVGRADGKRGSWT--DKTEARNEDEMYKILLNDYEQKQIL- 271  
DB 335 LTTKEKEAHTLQITLL-----AKENELRAFBEKLIAREGTEIQK-LIDD-----QKEVLG 383  
QY 272 -----MENAELKKVLQOMKXEMISLLSPK-----KKPRERAEDGTGTVASIDIED 317  
DB 384 SKMFELECEETIRKSLDKELQKBEELERQKVEIDHSEKLEKRNQ-----AMNKKFD 437  
QY 318 DSGELSRDSVWGLSCTVREQLTN-----SIRKQWRIL-KSHVEKLDNOAKSVHSEGLN 370  
DB 438 RYNEKEMDLEAKLTIKEREKIIQAEKELSLKQQLLSDKESLEDLQOEIKIRAEWTK 497  
QY 371 EEDVTSRQDHEQTEKLELEIERCKEMIKAAQOQLLOQQLATTCDDDTTSLLRDCVLLERK 430  
DB 498 KEEMI-----EBECKSLKKEEREYLRQLSELKSQIEKSRVHEEFLSKEVE-NLKQSK 551  
QY 431 ERLKEEWTL-----FKEQKNFER-----ERRSFTEAAR-----L 461  
DB 552 ERFEKEWELDKQAVYNKERRIRISEKEKFERQLLEGERLKKEBSALRVQIMQELDDI 611  
QY 462 GLERKAFERASWVKQQLNMTNFDHNS-----ENVKL 496  
DB 612 RLQRESFE-----ANMEHRSALQEKVKL 635

RESULT 12  
S74244  
serine/threonine-specific protein kinase (EC 2.7.1.1-) isoform I, Rho-associated - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 05-Oct-2004  
C:Accession: S74244  
R:Nakagawa, O.; Fujisawa, K.; Ishizaki, T.; Saito, Y.; Nakao, K.; Narumiya, S.  
FEBS Lett. 392, 189-193, 1996  
A:Title: ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein  
A:Reference number: S74244; MUID:96368048; PMID:8772201  
A:Accession: S74244  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1354 <NAK>  
A:Cross-references: UNIPROT:P70335; UNIPARC:UPI0000027020; EMBL:U58512; NID:gl514695; PI  
F:74-338/Domain: coiled coil; phosphotransferase; serine/threonine-specific protein kinase  
F:1229-1283/Domain: protein kinase C zinc-binding repeat homology <KIN>  
F:1229-1283/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 6.9%; Score 218.5; DB 2; Length 1354;  
Best Local Similarity 21.3%; Pred. No. 0.004;  
Matches 147; Conservative 123; Mismatches 229; Indels 191; Gaps 33;

QY 61 NTEQISYLDQBLTTGFP-SUYESKSKAEKREINIVAVLNCMNEILLVQRKNLLAQES 119  
DB 476 NLESASVQIEKEMLLQHRINEYQKVEQENKRNINEVSTLKDLEDLRK---ASQT 532  
QY 120 VETQNLKLGSDMDHLQSCYAKLKEQLETS---RREMIGLQERDRLOCKNRSLL---HQLL 173

Db 533 SOLANEKLTQLOKQEEANDLLRTESDTAVRLRKSHTEMSKISQLESNLRELQERNRIL 592  
QY 174 KNEKDEVQK---LQNIIASRATQYNHD-----VKRKEREYNKLERLHQL----- 215  
DB 593 ENSKSOADKYQLOQAVLEAERDRGDSMTGDLOARITSLOEEVVKHLKHNLERVEGER 652  
QY 216 -----VMNKDK---NIAMDVLNVYGRADGKRGSWRTDK---TEARNED----- 253  
DB 653 KEAQMDLHSEKEKNLEID-LNYKLKSIQORLEQEVNEHKVTKARLTDKHQSIEEAKSV 711  
QY 254 ---EMVKILLNDYER-----OKOILMENAEKLVLOOMKXEMISLLSPKQKPPR 300  
DB 712 ANCEMEKKLKEEREAREKAENRVVETEKQCSMLDVLKQ--SOQKLEHLT-----ENK 762  
QY 301 ERAEDGTGTVASIDIEDDSGE-----LSRDSVWGLSCTVRE----- 337  
DB 763 ERMEDEVKNLAL-QLQESNKRLLLONELKTOAFADNLKGLKQMKQKQINTLLEAKRLL 821  
QY 338 -----QLTNSIRKQWR-----ILKSHVEKLDNOAKSVHSEGL-- 369  
DB 822 EPELAQLT---KQYRGNEGQMRLEQDQLEAEQYFSTLYKTQVKELKEBIEEKNRENLRK 877  
QY 370 -----NEEDVISRQ---DHEQETKL-----ELEIERCKEMIKAAQOQLLOQQLA 410  
DB 878 IOELOSEKETLTQDLAETKAESBQLARGILEEQYFELTQESKKAASNRQBITDKDHT 937  
QY 411 TTCDDDTTSL--RDCVLL--EERKEL-----KEWTLPKEQ-----KKNFERRR 452  
DB 938 VSRLETSNVLTKDTEMLRKENEELNEMRTAAEEYKLLKKEEINNLKAAFEKNISTERT 997  
QY 453 STEAAIRLG--LERKAFEEERASWVKQQLNMTNFDHNSNVKLSAFSGSSDPDNL 510  
DB 998 LKTOAVNKLAETMNRKDFKIDRKK-----ANTQDLRKKEKENRKLOLELNQEREFNQ 1051  
QY 511 VHSRPROKHLHSV-ANGVPACTSKLTSLPASPSTSDPROTHSCVSEHSSISVLNITPPE 569  
DB 1052 VVK--HQELNDMQALVEECTHRLNEQMLQASKESDIEQLRA-----KLLOLSDST 1101  
QY 570 SKPSEVARETSQKWSQSRPSSR--EGCVS 598  
DB 1102 SVASFPSEADETD-----GNLPESRIEGWL 1126

RESULT 13  
T29095  
cardiac muscle factor 1 - chicken  
N:Alternate names: CMF1 protein  
C:Species: Gallus gallus (chicken)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T29095  
R:Wei, Y.; Bader, D.; Litvin, J.  
Development 122, 2779-2789, 1996  
A:Title: Identification of a novel cardiac-specific transcript critical for cardiac myo  
A:Reference number: 220564; MUID:96379739; PMID:8787752  
A:Accession: T29095  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: mRNA  
A:Residues: 1-1538 <WEI>  
A:Cross-references: UNIPROT:Q98940; UNIPARC:UPI000000FC28E; EMBL:U62026; NID:gl621106; P  
A:Experimental source: strain white leghorn; heart  
C:Function:  
A:Description: may be involved in cardiac myogenesis  
C:Keywords: cardiac muscle; heart

Query Match 6.9%; Score 218.5; DB 2; Length 1538;  
Best Local Similarity 21.8%; Pred. No. 0.0046;  
Matches 154; Conservative 112; Mismatches 242; Indels 197; Gaps 36;

QY 15 ENKNLSQYTSKSPSSLYSQOVLCSVPVLSK----- 47  
DB 558 ENSGIKE-----KLESASVSKQQLSCGVVSLGKEPEKISAKHNVNKTVLSNTTDLDDIE 612





A/Accession: S34842  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 323-381, 'HKAI', <GUE2>  
A/Cross-references: UNIPARC:UPI00017737D; EMBL:M28266  
A/Note: difference at carboxyl end due to frameshift error  
C/Comment: This protein is found as flocculent material in the parasitophorous vacuole.  
C/Superfamily: trichohyalin; calmodulin repeat homology  
C/Keywords: EF hand  
F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match 6.9%; Score 217.5; DB 2; Length 1909;  
Best Local Similarity 24.7%; Pred. No. 0.0067;  
Matches 125; Conservative 90; Mismatches 167; Indels 125; Gaps 25;

QY 26 TKMSPSSLYSQVLCSSVPLSKNVHGVFGVCTGENIEQISYLDQELTTTFGPPSLYERS 85  
DB 1188 TKASKETLQEQSDLEQERLAK-----EKLQEQSDLEQ-----ER 1223

QY 86 KSKEAKRELNIIVAVLNCMELLVLQRKNL-----LAQSVETQNLKLGSDMDHLQSCYAKL 141  
DB 1224 RAKEKLEQ-----QQSDLEQERLAKLEKLEQEQ-----SDLEQERRAKEKL 1263

QY 142 KEQ---LETSRREMIGLERDROLCKNRSLSHQL-----LKNKDEVOVKLQ----- 184  
DB 1264 QEQQSDLEQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKLEKLEQEQSDLEQ 1323

QY 185 -NIIASRATQYNHDKVRKEREYVKLKER---LHQLVMNKKDKNIAMDVLNYYVGRADGKRG 240  
DB 1324 ERLAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQERLAKLEKLEQEQSDLEQERRAKEKLQ 1383

QY 241 SWRTDKTEARNEDEMYKILLNDYEVYKQKQILMENAEKLVQOMKKEMISLLSPQKKKPR 300  
DB 1384 EQQSD---LEQDRLAKLEKLEQ---QQRDLQERRAKEKLQEQQSDL-----EQERRAK 1430

QY 301 ERAEDGTGTVAISDIEDDSGELSRDSVWGLSCDTVREQLTNSIRKQWIRLKSHVEKLDNQ 360  
DB 1431 EKLQEQQ-----SDLEQER-----RAKEKLEQ---QSDLEQERRAK---EKLQEQ 1470

QY 361 ASKVHSEGLNEEDVISRQ-DHEQE---TEKL-----ELEIE-RCKEMIKAQOQLLQOQILA 410  
DB 1471 QSDLEQERLAKLEKLEQQRDLQERRAKEKLQEQQSDLEQERRAKEKLQEQQSDLEQEQ--- 1528

QY 411 TTCDDDDTTSLLRDCYLLEKERLKEWTLPKQKKNFERRRS---FTEAAIRLGLERKA 467  
DB 1529 RLANEKLEQQRD---LEQERRAKEK---LQEQSDLEQERRAKEKLQEQQSDLEQERRA 1582

QY 468 FE--EERASWVKQOFLNMTNFDHONSE 492  
DB 1583 KEKLQEQQSDLEQERLAKLEKLEQQRD 1609

Search completed: June 12, 2006, 19:15:11  
Job time : 46 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:05:44 ; Search time 301 Seconds  
 (without alignments)  
 1889.982 Million cell updates/sec

Title: US-10-644-084-2  
 Perfect score: 3165  
 Sequence: 1 MGDWTVTPVLCTENKNLS.....CYSGCSSAPRSHGRDDLP 615

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : UniProt 7.2.\*  
 1: uniprot\_sprot.\*  
 2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Query	ID	Description
1	3165	100.0	615	1	ADIP MOUSE
2	3165	100.0	627	2	Q6ZV4 MOUSE
3	2937	92.8	613	1	ADIP RAT
4	2786.5	88.0	614	1	ADIP HUMAN
5	1558.5	49.2	545	2	Q6NRX3 XENLA
6	1554	49.1	554	2	Q6NRK1 XENLA
7	1355	42.8	576	2	Q6P942 BRARE
8	1148.5	36.3	495	2	Q4RWX4 TETNG
9	1142	36.1	528	2	Q4T607 TETNG
10	1096	34.6	480	2	Q4RU57 TETNG
11	362	11.4	398	2	Q6Z0V1 ORYSA
12	343	10.8	382	2	Q8GW47 ARATH
13	338	10.7	373	2	Q5AE7 ARATH
14	300	9.5	276	2	Q9FIE0 ARATH
15	260	8.2	1813	2	Q513P3 ENTHI
16	258	8.2	284	2	Q8GUK2 ARATH
17	251	7.9	1919	2	Q4RIP0 TETNG
18	243.5	7.7	1985	2	Q7RC59 PLAYO
19	240.5	7.6	1370	2	Q73732 XENLA
20	239	7.6	762	2	Q5OR43 ENTHI
21	238.5	7.5	1738	2	Q76329 DICDI
22	237	7.5	2332	2	Q22847 CAEEL
23	236	7.5	1024	2	Q54L07 DICDI
24	235	7.4	924	2	Q54107 dictyosteli
25	234.5	7.4	846	2	Q15738 dictyosteli
26	234.5	7.4	1679	1	MLP2 YEAST
27	233	7.4	987	2	Q8LI18 ORYSA
28	232.5	7.3	709	2	Q9GRG1 TETTH
29	231	7.3	1980	2	Q6FWE0 CANGA
30	229.5	7.3	1596	2	Q81J44 PLAF7
31	227	7.2	1738	2	Q51ED7 ENTHI

32	226.5	7.2	753	2	Q2SR10 MYCCA
33	226.5	7.2	798	2	Q5B8R8 EMENI
34	226.5	7.2	1175	2	Q8XNW6 CLOPE
35	226	7.1	1354	1	ROCK1 RABIT
36	226	7.1	2612	2	Q815X5 PLAF7
37	224.5	7.1	1723	2	Q4N897 THEPA
38	224	7.1	712	2	Q98SN5 CHICK
39	224	7.1	800	2	Q96X03 EMENI
40	224	7.1	2139	2	Q07569 ENTHI
41	223.5	7.1	3259	1	GOGB1 HUMAN
42	223	7.0	1699	2	Q4UHB3 THEAN
43	223	7.0	2042	2	Q6BUQ9 DEBHA
44	222.5	7.0	1558	2	Q4SEM9 TETNG
45	222	7.0	1946	2	Q97291 PLAF7

## ALIGNMENTS

RESULT 1  
 ADIP\_MOUSE  
 ID ADIP\_MOUSE STANDARD; PRT; 615 AA.  
 AC Q8VC66; Q8BG59; Q8C7X0; Q8K2F7;  
 DT 11-OCT-2004, integrated into UniProtKB/Swiss-Prot.  
 DT 01-MAR-2002, sequence version 1.  
 DT 07-FEB-2006, entry version 18.  
 DE Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-interacting protein).  
 GN Name=SBX2ip;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, SUBCELLULAR LOCATION, INTERACTIONS WITH AFIDIN AND ALPHA-ACTININ, AND TISSUE SPECIFICITY.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=22450629; PubMed=12446711; DOI=10.1074/jbc.M209832200;  
 RA Asada M., Irie K., Morimoto K., Yamada A., Ikeda W., Takeuchi M., Takai Y.,  
 RA "ADIP, a novel afadin- and alpha-actinin-binding protein localized at cell-cell adherens junctions.",  
 J. Biol. Chem. 278:4103-4111(2003).  
 [2]

NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA]  
 STRAIN=C57BL/6J; TISSUE=Brain cortex, Cerebellum, and Testis;  
 PubMed=16141072; DOI=10.1126/science.112014;  
 Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crome M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G., Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Humnietek L., Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuura S., Miki H., Mignone F., Miyake S., Morris K., Mottagui-Farzi S., Mulder N., Nakano N., Nakachi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada K., Shimada K., Silva D., Sinclair B.,

RA Sperling S., Stupka E., Sugura K., Sultana R., Takanaka Y., Taki K.,  
RA Tamoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brueic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawahima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami T., Waki K., Watanishi A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome."; [Science 309:1559-1563\(2005\)](http://www.sciencemag.org/content/309/5939/1559).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA]  
RC STRAIN=FBV/N; TISSUE=Mammary gland, and Salivary gland;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -|- FUNCTION: Belongs to an adhesion system which plays a role in the  
organization of homotypic, interneuronal and heterotypic cell-cell  
adherens junctions (AJs). May connect the nectin-afadin and E-  
cadherin-catenin system through alpha-actinin and may be involved  
in organization of the actin cytoskeleton at AJs through afadin  
and alpha-actinin.  
CC -|- SUBUNIT: Interacts with afadin and alpha-actinin.  
CC -|- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions  
(AJs). Not found at cell-matrix AJs.  
CC -|- TISSUE SPECIFICITY: Widely expressed.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AF532969; AA015015.1; -; mRNA.  
CC EMBL: AK049080; BAC33536.1; -; mRNA.  
CC EMBL: AK031356; BAC27363.1; -; mRNA.  
CC EMBL: AK043865; BAC31684.1; -; mRNA.  
CC EMBL: BC021749; AAH21749.1; -; mRNA.  
CC EMBL: BC031527; AAH31527.1; -; mRNA.  
CC Ensemble: ENSMUSG00000036825; Mus musculus.  
CC MGI: MGI:2139150; Ssx2ip.  
CC GO: GO:0005515; F:protein binding; IPI.  
CC Cell adhesion; Coiled coil.  
FT CHAIN 1 615 Afadin- and alpha-actinin-binding  
FT protein.  
FT /FTID=PRO\_0000064456.  
FT  
FT COILED 126 227 Potential.  
FT COILED 266 293 Potential.  
FT COILED 375 461 Potential.  
FT CONFLICT 47 47 K -> R (in Ref. 3; AAH31527).  
FT CONFLICT 406 406 Missing (in Ref. 2).  
FT CONFLICT 503 503 S -> SKPG (in Ref. 2; BAC33536).  
FT SEQUENCE 615 AA; 70956 MW; CC41D707859191F CRC64;  
SQ

Query Match 100.0%; Score 3165; DB 1; Length 615;  
Best Local Similarity 100.0%; Pred. No. 5e-133;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MGDWMTVDPVLCATENKNLSQYTTSETKMPSSLSYQQVLCSSVPLSKNVHGVGFCTGCE 60  
DB 1 MGDWMTVDPVLCATENKNLSQYTTSETKMPSSLSYQQVLCSSVPLSKNVHGVGFCTGCE 60  
QY 61 NIEOSISYLDDELTTFFPSPSYEESKSEAKRELINIVAVLNCMELLVLQKNLLAQESV 120  
DB 61 NIEOSISYLDDELTTFFPSPSYEESKSEAKRELINIVAVLNCMELLVLQKNLLAQESV 120  
QY 121 ETQNLKLGSDMDHLQSCYAKLKEQLETSRRREMIGLQERDRLQCKNRSLLHQLLNKDEK 180  
DB 121 ETQNLKLGSDMDHLQSCYAKLKEQLETSRRREMIGLQERDRLQCKNRSLLHQLLNKDEK 180  
QY 181 OKLQNIITASRATQYNDVVKREYKNLKERLHQLVMNKKDNITAMDVLNYVGRADGKRG 240  
DB 181 OKLQNIITASRATQYNDVVKREYKNLKERLHQLVMNKKDNITAMDVLNYVGRADGKRG 240  
QY 241 SWRTDKTEARNEDEMYKILLNDYEVROKQIILMENAELKKVLQOMKEMISLLSPKKKPR 300  
DB 241 SWRTDKTEARNEDEMYKILLNDYEVROKQIILMENAELKKVLQOMKEMISLLSPKKKPR 300  
QY 301 ERAEDGTGTVAISDIEDSGELSRDSVGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ 360  
DB 301 ERAEDGTGTVAISDIEDSGELSRDSVGLSCDTVREQLTNSIRKQWRLKSHVEKLDNQ 360  
QY 361 ASKVHSEGLNEEDVISQDHPQTEKLELETERCKEMIKAAQQLLQQLATTCDDDTTSL 420  
DB 361 ASKVHSEGLNEEDVISQDHPQTEKLELETERCKEMIKAAQQLLQQLATTCDDDTTSL 420  
QY 421 LRDCVLLSEERLKEEWTLPKEQKNPERERRSFTEAIRLGLERKAFEEERASWVKQOF 480  
DB 421 LRDCVLLSEERLKEEWTLPKEQKNPERERRSFTEAIRLGLERKAFEEERASWVKQOF 480  
QY 481 LNMNTFQHNSENKVLFSAFSGSSDPDNLI VHSRPRQKLSHVANGVPACTSKLTSLPA 540  
DB 481 LNMNTFQHNSENKVLFSAFSGSSDPDNLI VHSRPRQKLSHVANGVPACTSKLTSLPA 540  
QY 541 SPSTSDFRQTHSCVSEHSSISVLNITPEESKPESEVARESTQKWSVQSRPSREGCYGC 600  
DB 541 SPSTSDFRQTHSCVSEHSSISVLNITPEESKPESEVARESTQKWSVQSRPSREGCYGC 600  
QY 601 SSAPFSAHGDRDDLP 615  
DB 601 SSAPFSAHGDRDDLP 615  
RESULT 2  
Q69ZV4\_MOUSE PRELIMINARY; PRT; 627 AA.  
ID Q69ZV4\_MOUSE  
AC Q69ZV4\_MOUSE  
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2004, sequence version 1.  
DT 07-FEB-2006, entry version 10.  
DE MKIAA0923 protein (Fragment).  
GN Name=Ssx2ip; Synonyms=mKIAA0923;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Embryonic tail;  
RX PubMed=15368895; DOI=10.1093/dnares/11.3.205;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoaka S.,  
RA Soga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,  
RA Nagase T., Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
IV. The complete nucleotide sequences of 500 mouse KIAA-homologous  
cDNAs identified by screening of terminal sequences of cDNA clones

randomly sampled from size-fractionated libraries.";  
DNA Res. 11:205-218(2004).  
-----  
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
EMBL: AK173064; BAD32342.1; -: mRNA.  
DR Ensembl: ENSMUSG00000036825; Mus musculus.  
DR MGI: MGI:2139150; Ssx2ip.  
DR GO: GO:0005515; F:protein binding; IPI.  
FT NON\_TER 1  
SQ SEQUENCE 627 AA; 72053 MW; 7F2B9A2DF5F73E2D CRC64;  
  
Query Match 100.0%; Score 3165; DB 2; Length 627;  
Best Local Similarity 100.0%; Pred. No. 5.1e-133;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MGDWMTVTPVLCTENKNLSQYTSKMSPSLYSQVLCSSVPLSKNVHGVFGVFCG 60  
DB 13 MGDWMTVTPVLCTENKNLSQYTSKMSPSLYSQVLCSSVPLSKNVHGVFGVFCG 72  
  
QY 61 NIEQSYLDQELTTFGFSPSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV 120  
DB 73 NIEQSYLDQELTTFGFSPSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV 132  
  
QY 121 ETQNLKSGDMHQLQSCYAKLKEQLETSRREMIGLQERDROCKNRSIHLKNEKDEV 180  
DB 133 ETQNLKSGDMHQLQSCYAKLKEQLETSRREMIGLQERDROCKNRSIHLKNEKDEV 192  
  
QY 181 OKLQNIASRATQYNHDKVKKEREYNKLERLHOLVMNKKDKNIAMDLVNYVGRADGKRG 240  
DB 193 OKLQNIASRATQYNHDKVKKEREYNKLERLHOLVMNKKDKNIAMDLVNYVGRADGKRG 252  
  
QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLOQMKKEMISLLSPQKKPR 300  
DB 253 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLOQMKKEMISLLSPQKKPR 312  
  
QY 301 ERAEDGTGTVALSIEDDDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360  
DB 313 ERAEDGTGTVALSIEDDDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 372  
  
QY 361 ASKVHSEGLNEEDVTSRQDHQETEKLEIEIRCKEMIKAAQQLLQOQLATTCDDDTTSL 420  
DB 373 ASKVHSEGLNEEDVTSRQDHQETEKLEIEIRCKEMIKAAQQLLQOQLATTCDDDTTSL 432  
  
QY 421 LRDCYLLEKRLKEEWTLFKQKKNFERERSFTEAAILRLGLERKAFEEERASWVKQF 480  
DB 433 LRDCYLLEKRLKEEWTLFKQKKNFERERSFTEAAILRLGLERKAFEEERASWVKQF 492  
  
QY 481 LNMTFDHQNSENVKLFSAFGSSDDPNLIHVSRRPRQKKLHVSANGVPACTSKLTKSLPA 540  
DB 493 LNMTFDHQNSENVKLFSAFGSSDDPNLIHVSRRPRQKKLHVSANGVPACTSKLTKSLPA 552  
  
QY 541 SPSTSDFROTHSCVSHSHSISVLNITPESKPESEVARESTDOKWSVQSPRSPREGCYSC 600  
DB 553 SPSTSDFROTHSCVSHSHSISVLNITPESKPESEVARESTDOKWSVQSPRSPREGCYSC 612  
  
QY 601 SSAFRSAHGDRDDL 615  
DB 613 SSAFRSAHGDRDDL 627  
  
RESULT 3  
ADIP RAT STANDARD; PRT; 613 AA.  
AC Q8CG22;  
DT 11-OCT-2004, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAR-2005, sequence version 1.  
DE 07-FEB-2006, entry version 16.  
DE Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-interacting protein).  
GN Name=Sex2ip;  
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN 11  
RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, SUBCELLULAR LOCATION, AND  
RP INTERACTIONS WITH AFDIN AND ALPHA-ACTININ.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=22450629; PubMed=12446711; DOI=10.1074/jbc.M209832200;  
RA Asada M., Irie K., Morimoto K., Yamada A., Ikeda W., Takeuchi M.,  
RA Takai Y.;  
RT "ADIP, a novel afadin- and alpha-actinin-binding protein localized at  
RT cell-cell adherens junctions.";  
RL J. Biol. Chem. 278:4103-4111(2003).  
RN 12  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC TISSUE=Testis;  
RG NIH - Mammalian Gene Collection (MGC) project;  
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Belongs to an adhesion system, which plays a role in the  
CC organization of homotypic, interneuronal and heterotypic cell-cell  
CC adherens junctions (AJs). May connect the nectin-afadin and E-  
CC cadherin-catenin system through alpha-actinin and may be involved  
CC in organization of the actin cytoskeleton at AJs through afadin  
CC and alpha-actinin.  
CC -!- SUBUNIT: Interacts with afadin and alpha-actinin.  
CC -!- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions  
CC (AJs). Not found at cell-matrix AJs.  
CC -----  
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AF532970; AA015016.1; -: mRNA.  
DR EMBL: BC078687; AAH78687.1; -: mRNA.  
DR HSSP: P01096; 1HF9.  
DR Ensembl: ENSRNOG00000015425; Rattus norvegicus.  
DR RGD: 70849; Sex2ip.  
KW Cell adhesion; Coiled coil.  
FT CHAIN 1 613 Afadin- and alpha-actinin-binding  
FT protein.  
FT /FTID=PRO\_0000064457.  
FT COILED 126 227 Potential.  
FT COILED 266 293 Potential.  
FT COILED 375 461 Potential.  
SQ SEQUENCE 613 AA; 70679 MW; 55A6174A810AB8DE CRC64;  
  
Query Match 92.8%; Score 2937; DB 1; Length 613;  
Best Local Similarity 92.2%; Pred. No. 6.9e-123;  
Matches 567; Conservative 18; Mismatches 28; Indels 2; Gaps 1;  
  
QY 1 MGDWMTVTPVLCTENKNLSQYTSKMSPSLYSQVLCSSVPLSKNVHGVFGVFCG 60  
DB 1 MGDWMTVTPVLCTENKNLSQYTSKMSPSLYSQVLCSSVPLSKNVHGVFGVFCG 60  
  
QY 61 NIEQSYLDQELTTFGFSPSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV 120  
DB 61 NIEQSYLDQELTTFGFSPSYEESKSKAKRELNIIVAVLNCMNEELLVLQKNLLAQESV 120  
  
QY 121 ETQNLKSGDMHQLQSCYAKLKEQLETSRREMIGLQERDROCKNRSIHLKNEKDEV 180  
DB 121 ETQNLKSGDMHQLQSCYAKLKEQLETSRREMIGLQERDROCKNRSIHLKNEKDEV 180  
  
QY 181 OKLQNIASRATQYNHDKVKKEREYNKLERLHOLVMNKKDKNIAMDLVNYVGRADGKRG 240  
DB 181 OKLQNIASRATQYNHDKVKKEREYNKLERLHOLVMNKKDKNIAMDLVNYVGRADGKRG 240  
  
QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLOQMKKEMISLLSPQKKPR 300  
DB 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLOQMKKEMISLLSPQKKPR 300  
  
QY 301 ERAEDGTGTVALSIEDDDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360  
DB 301 ERAEDGTGTVALSIEDDDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360

```
QY 361 ASKVHSEGLNEBDVISRODHEQTEKLEIEIRCKEMIKAOQLOQLATTCCDDTTSL 420
DB 361 ASKVHSEGHEDVDVISRODHEQTEKLEIEIRCKEMIKAOQLOQLATTCCDDTTSL 420
QY 421 LRDCYLLEKRLKEWTLFKQKNFERRSFTEAAIRLGLERKAFEEERASVVKQOF 480
DB 421 LRDCYLLEKRLKEWTLFKQKNFERRSFTEAAIRLGLERKAFEEERASVVKQOF 480
QY 481 LNMTHFDHONSENVKLFSAFSSSDPDNLIHVSRRPROKXKLVANGVACTSKLTKSLPA 540
DB 481 LNMTHFDHONSENVKLFSAFSSSDPDNLIHVSRRPROKXKLVANGVACTSKLTKSLPT 540
QY 541 SPSTDFRTHSCVSEHSHSISVLNITPESKSEVARESTDQKWSVQSRPSSREGCYGCG 600
DB 541 SP-SDFCPSRCVSEHSPVSLVTPETKNEVGRESTDQKWSVQSRPSSREGCYGCG 598
QY 601 SNAFRSAHGDRDLDP 615
DB 599 SSAYTSSHVERDLDP 613

RESULT 4
ADIP HUMAN STANDARD; PRT; 614 AA.
ID ADIP HUMAN Q9Y2D8; O6UL61; Q7L168; Q9UIX0;
AC Q9Y2D8; O6P2P8; Q6UL61; Q7L168; Q9UIX0;
DT 11-OCT-2004, integrated into UniProtKB/Swiss-Prot.
DT 11-OCT-2004, sequence version 3.
DE Q9Y2D8-2006, entry version 23.
DE Afadin- and alpha-actinin-binding protein (ADIP) (Afadin DIL domain-
interacting protein) (SSX2-interacting protein).
GN Name=SSX2IP; Synonyms=KIAA0923;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], INTERACTIONS WITH SSX2 AND SSX3, AND
RP SUBCELLULAR LOCATION.
RX PubMed=12007189; DOI=10.1002/gcc.10073;
RA de Bruijn D.R.H., dos Santos N.R., Kater-Baats E., Thijssen J.,
RA van den Berk L., Stap J., Baalman M., Schepens M., Merckx G.,
RA van Kessel A.G.;
RT "The cancer-related protein SSX2 interacts with the human homologue of
a Ras-like GTPase interactor, RAB3IP, and a novel nuclear protein,
RT SSX2IP."
RL Genes Chromosomes Cancer 34:285-298(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 1).
RC TISSUE=Testis;
RA Lu L., Huang X.Y., Xu M., Yin L.L., Li J.M., Zhou Z.M., Sha J.H.;
RT "Cloning a new transcript of X breakpoint 2 interacting protein
RT (SSX2IP) in testis."
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032; DOI=10.1093/dnares/6.1.63;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hiroseawa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM 1).
RC TISSUE=Testis;
RG The German cDNA consortium;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 2).
```

```
RC TISSUE=Brain, and Eye;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzly D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting M., Madan A., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Belongs to an adhesion system, which plays a role in the
organization of homotypic, interneuronal and heterotypic cell-cell
adherens junctions (AJs). May connect the nectin-afadin and E-
cadherin-catenin system through alpha-actinin and may be involved
in organization of the actin cytoskeleton at AJs through afadin
and alpha-actinin (By similarity).
CC -!- SUBUNIT: Interacts with afadin and alpha-actinin (By similarity).
Interacts with SSX2 and SSX3. Does not interact with SSX1 and
SSX4.
CC -!- SUBCELLULAR LOCATION: Localized at cell-cell adherens junctions
(AJs). Not found at cell-matrix AJs (By similarity). Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9Y2D8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9Y2D8-2; Sequence=VSP_011724, VSP_011725;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Widely expressed, with the highest expression
in brain, intermediate expression in kidney, testis, spinal cord,
liver, heart, lung, skeletal muscle, ovary, fetal liver and fetal
brain, and little to no expression in pancreas and spleen. All
specific brain regions showed intermediate to high expression,
with highest expression in amygdala. Also expressed in fetal
tissues, mainly in liver and brain.
CC -!- DOMAIN: Both the N-terminal (up to position 79) and the C-terminal
(from position 304) sequences are required for interaction with
SSX2.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY367055; AAQ72373.1; -; mRNA.
DR EMBL; AB023140; BAA76767.2; ALT_INIT; mRNA.
DR EMBL; AL133046; CAB61373.1; -; mRNA.
DR EMBL; BC033637; AAH33637.1; -; mRNA.
DR EMBL; BC064389; AAH64389.1; -; mRNA.
DR PIR; T42649; T42649.
DR Ensembl; ENSG00000117155; Homo sapiens.
DR HGNC; HGNC:16509; SSX2IP.
DR MIM; 608690; gene.
DR LinkHub; Q9Y2D8; -.
KW Alternative splicing; Cell adhesion; Coiled coil; Nuclear protein.
FT CHAIN 1 614 Afadin- and alpha-actinin-binding
FT protein.
FT /FTid=PRO_0000064455.
FT COILED 131 227 Potential.
FT COILED 266 293 Potential.
FT COILED 374 460 Potential.
FT VARSPPLIC 557 560 SSIN -> RAKN (in isoform 2).
FT /FTid=VSP_011724.
```

```
FT VARSPLIC 561 614 Missing (in isoform 2).
FT CONFLICT 232 232 /FtId=VSP_011725.
FT CONFLICT 405 405 V -> A (in Ref. 2).
FT CONFLICT 456 456 Missing (in Ref. 5; AAH64389).
FT CONFLICT 576 576 G -> V (in Ref. 2).
FT CONFLICT 578 578 C -> R (in Ref. 4).
SQ SEQUENCE 614 AA; 71236 MW; 427903BF86A6FE31 CRC64;

Query Match 88.0%; Score 2786.5; DB 1; Length 614;
Best Local Similarity 87.5%; Pred. No. 3.5e-116;
Matches 538; Conservative 35; Mismatches 41; Indels 1; Gaps 1;

QY 1 MGDWMTVTDVLTCTENKLSOYVSETKMPSSLYSQOVLCSVPLSKNVHGVGVCTGE 60
DB 1 MGDWMTVTDVLTCTENKLSOYVSETKMPSSLYSQOVLCSVPLSKNVHGVGVCTGE 60

QY 61 NIEQISYLDQELTTFGPPSLYBESKSKAKRELNIIVAVLNCNELLVLQKNLLAQESV 120
DB 61 NIEQISYLDQELTTFGPPSLYBESKSKAKRELNIIVAVLNCNELLVLQKNLLAQESV 120

QY 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSLLHLLKNEKDEV 180
DB 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSLLHLLKNEKDEV 180

QY 181 OKLQNIIASRATQYNHDKRERKREYNKLERHOLVNMKKDKNIAMDVLNYVGRADCKRG 240
DB 181 OKLQNIIASRATQYNHDKRERKREYNKLERHOLVNMKKDKNIAMDVLNYVGRADCKRG 240

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300
DB 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300

QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360
DB 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360

QY 361 ASKVHSEGLNEEDVLSRODHEQTEKLELEIERCKEMIKAQOOLQOQLATTCDDDTTSL 420
DB 361 ASKVHSEGLNEEDVLSRODHEQTEKLELEIERCKEMIKAQOOLQOQLATTCDDDTTSL 420

QY 421 LRDCYLLEEKERLKEBWTFLFKEQKNFERERSFTEAAILGLERKAFEEERASWVKQF 480
DB 421 LRDCYLLEEKERLKEBWTFLFKEQKNFERERSFTEAAILGLERKAFEEERASWVKQF 480

QY 481 LNMTPDHQNSENVKLFSAFSGSDPDNLIVHSRPRQKHLVANGVAPACTSKLTKSLPA 540
DB 481 LNMTPDHQNSENVKLFSAFSGSDPDNLIVHSRPRQKHLVANGVAPACTSKLTKSLPA 540

QY 541 SPSTDFFRTHSCVSHSHSISVLNITPEESKPEVARESTDQKWSVQSPSSREGCYSGC 600
DB 541 SPSTDFFRTHSCVSHSHSISVLNITPEESKPEVARESTDQKWSVQSPSSREGCYSGC 600

QY 601 SSAFRSAHGDRDDLP 615
DB 601 SLSYTNSHVEKDDLP 614

RESULT 5
Q6NRX3 XENLA PRELIMINARY; PRT; 545 AA.
AC Q6NRX3_
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE MGC81138 protein.
GN Name=MGC81138;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Murthy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Giummo J.J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC070587; AAH70587.1; -; mRNA.
SQ SEQUENCE 545 AA; 63836 MW; 79A2CB644DBA2CA3 CRC64;

Query Match 49.2%; Score 1558.5; DB 2; Length 545;
Best Local Similarity 55.7%; Pred. No. 1.3e-61;
Matches 326; Conservative 102; Mismatches 112; Indels 45; Gaps 13;

QY 1 MGDWMTVTDVLTCTENKLSOYVSETKMPSSLYSQOVLCSVPLSKNVHGVGVCTGE 60
DB 1 MGDWMTVTDVLTCTENKLSOYVSETKMPSSLYSQOVLCSVPLSKNVHGVGVCTGE 60

QY 61 NIEQISYLDQELTTFGPPSLYBESKSKAKRELNIIVAVLNCNELLVLQKNLLAQESV 120
DB 61 NIEQISYLDQELTTFGPPSLYBESKSKAKRELNIIVAVLNCNELLVLQKNLLAQESV 120

QY 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSLLHLLKNEKDEV 180
DB 121 ETQNLKLGSDMDHLQSCYAKLEQLETSRREMIGLQERDRQLQCKNRSLLHLLKNEKDEV 180

QY 181 OKLQNIIASRATQYNHDKRERKREYNKLERHOLVNMKKDKNIAMDVLNYVGRADCKRG 240
DB 181 OKLQNIIASRATQYNHDKRERKREYNKLERHOLVNMKKDKNIAMDVLNYVGRADCKRG 240

QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300
DB 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300

QY 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360
DB 301 ERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKQWRILKSHVEKLDNQ 360

QY 361 ASKVHSEGLNEEDVLSRODHEQTEKLELEIERCKEMIKAQOOLQOQLATTCDDDTTSL 420
DB 361 ASKVHSEGLNEEDVLSRODHEQTEKLELEIERCKEMIKAQOOLQOQLATTCDDDTTSL 420
```

```
Db 349 ATP-----DENGMIARGDHEQELGLKLINEIQCKETIKIQOQLLKQFVSP-RDDTSTL 401
Qy 421 LDCYLLBEKERLKEBWTLFKEQKNFERERSFTEAAIRLGLERKAFEEERASVVKQOF 480
Db 402 LQDCYLLDEKRLQEBWLFNQKNFKERKNFTAAIRLGHKAFEEEDRAAWLKHQF 461
Qy 481 LNMNMF-DHONSENVKLFSAFGSSDPDNLIVHSPROKHLHSHVANGVPACTSKLTSLP 539
Db 462 LNMVTFDHNKEE-KRANGVHFSPQDCHRLHSRTHDRHLASSGDHY-----QRKPTLP 516
Qy 540 ASPSTSDFRQTHSCVSEHSSISVLNITPES---KPSEVARESTD 581
Db 517 ITPS-----SKHS-----LTQESVAVWRDSSISPNGTD 544

RESULT 6
Q6NRK1_XENLA
AC Q6NRK1_XENLA PRELIMINARY; PRT; 554 AA.
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 10.
DE MGC83757 protein.
GN Name=MGC83757;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8335;
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Oocytes;
RA Klein S., Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted under the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC070749; AAH70749.1; -; mRNA.
SQ SEQUENCE 554 AA; 64680 MW; 439EA3ABE3C2C98D CRC64;
```

```
Query Match 49.1%; Score 1554; DB 2; Length 554;
Best Local Similarity 55.2%; Pred. No. 2.2e-61;
Matches 321; Conservative 107; Mismatches 122; Indels 32; Gaps 10;

Qy 1 MGDMMVTDPVLCTENKNLSQVTSQTKMSPSLXSQQVLCSSVPLSKNVHGVGVGFCTGEE 60
Db 1 MGDRTTSLP---ESDKILOYSCEIRMSPTSLPS----PSHVSANNLSGVSVYTFCTED 52
Qy 61 NIEQSYLDDELTTTFPPSLYEESKKEAKRELNIIVAVLNCMELLVLQRKLLAQSRV 120
Db 53 NLEQCITYIDQLRTIGFPTQAVSKNGEG-RKLHLVSIIINCIYELLQNSQTKMSNEEV 111
Qy 121 ETQNLKLGSDMDHLQSCYAKLKEOLETSRRRMIQLOERDQLOQCNRSIHLHLLKNEDEV 180
Db 112 ETQLKINGDLEYLQSHQROKQLEATKRENCALQERDRQCNKRNLLQLLKNKEEV 171
Qy 181 QKLQNIITASRATQYNHDKRERENYKLKERLHQLVMNKKDKNIAMDVLNVYGRADKRG 240
Db 172 QKLQNIIASRSTQYNHVSVKRERENYKLKERLYQLVMDKDKISIDVLNVYGRADKRS 231
Qy 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300
Db 232 SWRTGKTDAKNEBMYKVLNDYEQKQILMENAELKKVLQOMKEMISIVS--QRKTK 289
Qy 301 BRAEDGTGCTVAISDIEDDSGSLSDSVNGLSCDVTREOLTNSIRKOWRLKSHVEKLDNQ 360
Db 290 EKLEDSGTGTVT-SDIEEIIADSKENLSCEAVREQLISSIRQOWRLKSHMEKLDNQ 348
Qy 361 ASKVHSEGLNEEDVISRDHEQETEKLEIERCKEMIKAAQQLLQOQLATTTCDTDTSL 420
Db 349 ACLNVPTPDENGLIARAEHEQELDKLISEIQCKETIRSQQLLKQSLVPRDDTSLK 408
Qy 421 LRDCYLLBEKRLKEBWTLFKEQKNFERERSFTEAAIRLGLERKAFEEERASVVKQOF 480
Db 409 LQDCYLLDEKRLQEBWLFNQKNFKERKNFTAAIRLGHKAFEEEDRAAWLKHQF 468
Qy 481 LNMNMF-DHONSENVKLFSAFGSSDPDNLIVHSPROKHLHSHVANGVPACTSKLTSLP 539
Db 469 LNMVTFDHNKEERKV-PGVHFSQDNCRLHSPHDPKVLASSGD-----YSRRPSKALP 523
Qy 540 ASPSTSDFRQTHSCVSEHSSISVLNITPESKPESEVARESTD 581
Db 524 ITSS-----SKHSLTQIESIS---WRDSSISPNSTD 551

RESULT 7
Q6P942_BRARE
ID Q6P942_BRARE PRELIMINARY; PRT; 576 AA.
AC Q6P942;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Hypothetical protein zgc:73314.
GN ORFNames=zgc:73314;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Wild-type; TISSUE=Eye;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
```



[illegible]

ID	Q4RWX4_TETNG	PRELIMINARY;	PRT;	495 AA.
AC	Q4RWX4;			
DC	19-JUL-2005,	integrated into UniProtKB/TrEMBL.		
DT	19-JUL-2005,	sequence version 1.		
DT	07-FEB-2006,	entry version 3.		
DE	Chromosome 15 SCAR14981,	whole genome shotgun sequence. (Fragment).		
GN	ORFNames=GSTENG00027654001;			
OS	Tetraodon nigroviridis (Green puffer).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;			
OC	Tetraodontidae; Tetraodontidae; Tetraodon.			
OX	NCBI_TaxID=99883;			
RN	[1]			
RN	NUCLEOTIDE SEQUENCE.			
RX	PubMed=15496914; DOI=10.1038/nature031025;			
RA	Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,			
RA	Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,			
RA	Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,			
RA	Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,			
RA	Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,			
RA	Blemond C., Skalli Z., Catrocilo L., Poulain J., De Berardinis V.,			
RA	Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,			
RA	Parra G., Lardier G., Chappe C., McKernan K.J., McEwan P., Bosak S.,			
RA	Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,			
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,			
RA	Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,			
RA	Winkler P., Lander E.S., Weissenbach J., Roest Crolius H.;			
RT	"Genome duplication in the teleost fish Tetraodon nigroviridis reveals			
RT	the early vertebrate proto-karyotype."			
RL	Nature 431:946-957 (2004).			
RN	[2]			
RN	NUCLEOTIDE SEQUENCE.			
RG	Genoscope; Whitehead Institute Centre for Genome Research;			
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.			
CC	-!- CAUTION: The sequence shown here is derived from an			
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is			
CC	preliminary data.			
CC	-----			
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>			
CC	Distributed under the Creative Commons Attribution-NoDerivs License			
CC	-----			
DR	EMBL; CAEO1014981; CAG07108.1; -; Genomic_DNA.			
FT	NON TER			
FT	1			
SQ	SEQUENCE 495 AA; 56767 MW; 8DB4FE914438457A CRC64;			
	Query Match 36.3%; Score 1148.5; DB 2; Length 495;			
	Best Local Similarity 46.9%; Pred. No. 2.le-43;			
	Matches 253; Conservative 83; Mismatches 139; Indels 65; Gaps 9			
QY	24 SETKMSPLSLYSQQVLCSSVPLSKNVHGVCFCTGENIFQSISYLDOLBTLTFGF--PSL 81			
Db	1 SHLTWSPTRPNVNWPLYPFLPLSKNSXSVISAFCTEDNIPQCHISYMNQELSSLGFSPPSI 60			
QY	82 YEESKSKAKRELNIIVALNCWELLVQRKNLLAQESVETQNLKLGSDMDHLQSCYAKL 141			
Db	61 GHTSPGNFG--LNTVAALNVYELVQIHRRSMCTLEBKDKHLKXSSLEHMQSSNRL 117			
QY	142 KEQLETS--RREMIGLQERDROLQCKNRSLLHQLLKNEKDEVQKLNIIASRATQYNHDKVR 200			
Db	118 KARRSSSGTTKFGLIHETROLQLKI TLQSLCKTEKDEVQKLQNIASRATQYSHDAKR 177			
QY	201 KEREYNKLKERLHOLVWNKKDKNTAMDVNVYGRADKRGSSWTDKTEARNEDEMYKILL 260			
Db	178 KEREAAKLKERLSQLLIDRKEKLAIDVNLGRADKRGKSHWKTAGTASHEGEMYSKLL 237			
QY	261 NDYEYRQKILMENAELKVLQOMKEMISLLSPQKKKPRERAEDGTGTVAISDIEDDGS 320			
Db	238 SDYEASQRSLVENSELKKVLQOMKRDVHILSP--RQAREPRADDQSQBPASSDGEKGP 295			
QY	321 ELSRDSVWMLSCDTVREQLTNSIRKQWRILKSHVEKLDNQASKVHSEGLNEEDVISRQDH 380			
Db	296 DCSREAL--DOSCEHAREOLANSVROOKLRSHVEKLDFOAFQVSOLOAKPKDVI PREAH 354			

RESULT 8  
Q4RWX4 T

```

QY 381 EOTEKLEIEIRCKEMIKAOQQLQQQLATTCDTDTSLLRDCYLLBEKRLKEEWTLF 440
DB 355 EDEMREVRVOCKEFIOAQOQLQQQLNSLDDDTAFLHDCYTLLEKRLKEEWTLF 414
QY 441 BQKQNFERRRSFTEAARLGLERKAFEBEERASWVKQOFLNMTNFDHQNSENVKLFSAF 500
DB 415 BQKQNFERRRSFTEAARLGLRGRD-----VRQSAL----- 444
QY 501 SGSSPDNLVHSRPROQKLHVSANGVPACTSKLTLSPASSTSDFTROTHSCVSEHSI 560
DB 445 -----PCARLIE-----CESLEBKGLPGGPSC-----LAESVSEHDSV 478

RESULT 9
Q4T607_TETNG
ID Q4T607_TETNG PRELIMINARY; PRT; 528 AA.
AC Q4T607;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Chromosome undetermined SCAF9008, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0006544001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAEA01090008; CAF1675.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 528
SQ SEQUENCE 528 AA; 60582 MW; EA51CD5B7F4AD26 CRC64;

Query Match 36.1%; Score 1142; DB 2; Length 528;
Best Local Similarity 45.8%; Pred. No. 4.4e-43;
Matches 248; Conservative 96; Mismatches 111; Indels 86; Gaps 9;

QY 15 ENKNIQSQTSETKMSPSSLSQOVLCSVPPLSKNVHGVFGVFCENTIEQ-----S 65
DB 4 EYKDVCGSPSECTPMPRFSQ-----SSLPLQHTSYILSTFTCTHNVQECLLRSQITS 59
QY 66 ISYL--DQELATTGFPFLSYESKSKARELNIVAVLNCMNBELLVLRKNLLAQESVETQ 123

```

```

DB 60 MSCLCVTCQAASGLPPVWPEPGSGS---EMNVAVLNCMDHLIQLYHRSLSRSLSELE 116
QY 124 NLKLGSDMDHLQSCYAKLKEQLETSRRRMIQLOERDROLQKNRSLHOLLKNEKEVQKL 183
DB 117 QHKSSSHVDYLQLTSAKLQELTSKRNTGLLERERLQKAKSLQKSLKNEKEVQKL 176
QY 184 QNIATSRATQYNDHVKREKREYNKRLKRLHOLVMNKDKNIAMDVLNVVGRADGRGSGWR 243
DB 177 QNIATSRASQYNHMKRERFNKRLNQLLSDKREKQKQIDVLNIGRADGRSLWK 236
QY 244 TDKTEARNEDEMYKILLNDYEQKQILMENAELKVKVLOQMCKEMISLLSPQKKPRERA 303
DB 237 TDKTEARHEGQLFKTLTSLSDYELRQRELLLENAELNKVLQOMKGEMTSVLA--SNKSTLTG 294
QY 304 EDGTGTVAISDIEDSGELSDSVGLSCDVTTRQOLTNIRKQWRLILKSHVKEKLDNQASK 363
DB 295 DDGV-TQAKSEDEBEVDFSSKESV-ELFCVHAREKLTNSVRLQWKLNHVERLDSQAS- 351
QY 364 VHSEGLNEEDVISRQDHEQTEKLEIERCKEMIKAOQQLL----- 405
DB 352 -----LNNADVVSRETHEEVDKLTETLQOQFDLIQTQQQLLHVRLKLVQESSQGLGN 406
QY 406 -----OOQLATTCDDDTTSLLRDCY 425
DB 407 HICHTCGICLLRCSDLIVCSEALDGAAGAAAANPFIVSQQQLSSQCDASSVHGSCP 466
QY 426 LLEEKERLKEEWTLFKQKQKNFERERRSFTEAARLGLERKAFEBEERASWVKQOFLNMTN 485
DB 467 MLQEKESLREEMKVLBEQKIFERRRNFTEAARLSDERKSFEEDRAWLKHQFLNLS 526
QY 486 F 486
DB 527 F 527

RESULT 10
Q4RU57_TETNG
ID Q4RU57_TETNG PRELIMINARY; PRT; 480 AA.
AC Q4RU57;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 3.
DE Chromosome 1 SCAF14995, whole genome shotgun sequence. (Fragment).
GN ORFNames=GSTENG00028943001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RL the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

```

CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: CAAE01014995; CAG08075.1; -, Genomic\_DNA.  
 CC FT NON\_TER 1  
 CC SQ SEQUENCE 480 AA; 55095 MW; 52F347B6F8229531 CRC64;  
 CC  
 CC Query Match 34.6%; Score 1096; DB 2; Length 480;  
 CC Best Local Similarity 50.1%; Pred. No. 4.5e-41;  
 CC Matches 237; Conservative 91; Mismatches 111; Indels 34; Gaps 9;  
 CC -----  
 CC  
 CC 15 ENKULSQYTSKMSPSLYSQVLCSSVPLSKNVHGVFGVCTGNIHQ-----S 65  
 CC 4 EVKDVCGPSSEKTPMPRPFQ-----SSLPQHRTSYILSTFTCHNVQECILRISQITS 59  
 CC 66 ISVL--DQELTTFGSPSLYEESKSEAKRELINIVAVLNCMNEILLVLRKNLLAQESVETQ 123  
 CC 60 MSCLCVTAASLGLPPVWPEPGSGS---EMNVAVLNCMHDLIQLYHRSLSLESLELE 116  
 CC 124 NLKLGSDMHLQSCYAKLKEQLETSRRMIGLQERDQLOCKNRSILHQLKNEKDEVKQL 183  
 CC 117 QHKSSHDVYLQLTSLRKEQLEISKRENTGLRERRRLQAKSLQHSKNEKEVEVQL 176  
 CC 184 QNIIASRATQYNHDKREREYNKLERLHQLVMNKKXNIAMDVLNYYVGRADGRGWSR 243  
 CC 177 QNIIASRASQYNHEMKREREFNKLKERLNLQSLDKREKKAIDVLNLSIGRADGRKSLWK 236  
 CC 244 TDKTEARNEDEMYKILLNDYERQKQILMENAEKLVLOQMKEMISLLSPQKKPRERA 303  
 CC 237 TDKTEAKHEGQFLKLLSYELRQRELLLENALNKVLQOMKGMTSVLA--SNKSTLTG 294  
 CC 304 EDGTGTVALSDIEDSGELSRDSVWGLSCDTVREQLTNSIRKOWRLKSHVEKLDNQASK 363  
 CC 295 DGVG-TQAKSEDEEVDSDKESV-ELFCVHAREKLTNSVRLQWRKLNHVERLDSQAS- 351  
 CC 364 VHSEGLNEEDVTSRDHEQTEKLELEIERCKEMIKAAQQLLQQLLQQLATTCDDDTTSLRD 423  
 CC 352 -----LNNADVVSRETHEBEVDKLTETIQDFKDLIQTOQLLHQLQSSQCDDEASSVHGS 406  
 CC 424 CYLLEEKERLKEEWLFLKQKKNFRERSFTFAAIRGLG-----RKAFEE 470  
 CC 407 CPMLOKESLREBWKVLEQKIFERERNFTFAAIRLSDEVIFFPAPRPAGE 459  
 CC  
 CC RESULT 11  
 CC Q620V1 ORYSA  
 CC ID Q620V1\_ORYSA PRELIMINARY; PRT; 398 AA.  
 CC AC Q620V1;  
 CC DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
 CC DT 05-JUL-2004, sequence version 1.  
 CC DT 07-FEB-2006, entry version 5.  
 CC DE Putative synovial sarcoma, X breakpoint 2 interacting protein.  
 CC GN Name=OSUNB0070J06.4;  
 CC OS Oryza sativa (japonica cultivar-group).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
 CC OC Ehrhartoideae; Oryzaceae; Oryza.  
 CC OX NCBI\_TaxID=39947;  
 CC RN [1]  
 CC RP NUCLEOTIDE SEQUENCE.  
 CC RA Sasaki T., Matsumoto T., Katayose Y.;  
 CC RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AP005478; BAD05632.1; -, Genomic\_DNA.  
 CC DR Gramene; Q620V1; -;  
 CC SQ SEQUENCE 398 AA; 44595 MW; 6E401C5AAB8C17FA4 CRC64;  
 CC

Query Match 11.4%; Score 362; DB 2; Length 398;  
 Best Local Similarity 26.1%; Pred. No. 1.7e-08;  
 Matches 104; Conservative 95; Mismatches 154; Indels 46; Gaps 12;  
 -----  
 QY 28 MSPSSLYSQVLCSSVP-----LSKNVHGVFGVCTGNIHQESISYLDQELT 74  
 DB 1 MSASRF--DLRASSPHQHAIGGGGGGGGSDGGGAAAAAFAADAGNLBHCARYLNQTLV 58  
 QY 75 TFGPSPSLYEESKSEAKRELINIVAVLNCMNEILLVLRKNLLAQESVETQNLKLGSDMDHL 134  
 DB 59 TFGFPA-----SLDLFATDPVSIARTCNCIYALLOQRDIEFRESTNDLRQMQSDISRL 114  
 QY 135 QSCYAKLKEQLETSRRMIGLQERDQLOCKNRSILHQLKNEKDEVKQLNIIASR--AT 192  
 DB 115 EAKIERMDAQLAAKRELATLRTTEAKNTATLKAQIDKLQERDERFQKN--VIGNQVRT 172  
 QY 193 QYNHVDKREKEREYNKLERLHQLVMNKKDN--IAMDVLNYYVGRADGRGWSRTDKTEAR 250  
 DB 173 QOIHEMKKREKEYIKLQEKLNQVLMKKESRSRSGMEINMLLOKGRQGRGTWSGKK----- 228  
 QY 251 NEDEMYKILLNDYERQKQILMENAEKLVLOQMKEMISLLSPQKKPRERAEDGTGTV 310  
 DB 229 NDNDYYKMLVDAYEVKKQELMQENADLRALLRSMQMDMREFLNAPNVP-QPAVVGNQGR 287  
 QY 311 AISDIEDSGELSRDSVWGLSCDTVREQLTNSIRKOWRLKSHVEKLDN--QASKVHSEG 368  
 DB 288 EAGSQSPSLG--GKTDVFDLPFMARDQIEESLRTKMASIKARMTQLQDAQKGAEVTSEA 345  
 QY 369 LNEEDVISRDHEQTEKLELEIERCKEMIKAAQQLLQ 407  
 DB 346 T-----ERELE-LEAQLVEARSIIQEQASLMSK 372  
 -----  
 RESULT 12  
 Q620V1 ARATH  
 ID Q620V1\_ARATH PRELIMINARY; PRT; 382 AA.  
 AC Q620V1;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 9.  
 DE Hypothetical protein (At2g18870).  
 GN OrderedlocusNames=At2g18870;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Bowser L., Carninci P.,  
 RA Chan M.M., Chang C.H., Dale J.M., Hayashizaki Y., Hsuan V.W.,  
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,  
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
 RA Tang C.C., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G., Yuan S.,  
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AK119090; BAC43664.1; -, mRNA.  
 CC DR HSSP; P05412; 1FOS.  
 CC DR TAIR; At2g18876; -;  
 CC KW Hypothetical protein.



```
Db 186 QKGRQGTWNGKKTDT---DFYKIVDAYEAKNQELMAENTSLRALLRSMQTDMDRFL 241
QY 293 SPOKKPRERAEDGTGTVAISD 314
Db 242 N-----APNGSATLAGE 254

RESULT 15
ID Q513P3_ENTHI PRELIMINARY; PRT; 1813 AA.
AC Q513P3;
DT 07-JUN-2005, integrated into UniProtKB/TrEMBL.
DT 07-JUN-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Viral A-type inclusion protein repeat, putative.
GN ORFNames=82.t00025;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B.J., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Church C.M., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K.L., Ormond D., Squares R.,
RA Whitehead S., Quail M.A., Rabinowitsch E., Norbertczak H., Price C.,
RA Wang Z., Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S.,
RA Lohia A., Foster P.G., Sicheritz-Ponten T., Weber C., Singh U.,
RA Mukherjee C., El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M.,
RA Barrell B.G., Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAFB01000312; EAL48061.1; -; Genomic_DNA.
CC InterPro; IPR002017; Spectrin.
CC Pfam; PF02524; KID; 2.
CC Coiled coil.
CC SEQUENCE 1813 AA; 214154 MW; 51D9CE79257621FB CRC64;

Query Match 8.2%; Score 260; DB 2; Length 1813;
Best Local Similarity 19.3%; Pred. No. 0.0029;
Matches 121; Conservative 114; Mismatches 188; Indels 204; Gaps 24;

QY 11 VLCTENKLSQYTSYTKMSPSSLYSQVLCSYVPLSNVHGVGVCTGENTEQSISYLD 70
Db 452 IICDNNKEIAFKBEQ-----ENLQKELNQIK 478

QY 71 Q-----ELTTGFPSPLYEESKSKAKREL--NIVAVLNCMVELLVQKRLLAQB 118
Db 479 EEKQKTENKVELVDVYTKQENELNKLKEEKEIFNEKTTIENSINQ--IVEEKNKLTEE 536

QY 119 -----SVETQNL-----KLGSMDHLSQCYAKLKEQLTSRR----- 150
Db 537 KESIKQELDSIKADNSTKELEINKINEEKQNQLNDYDTVQOEKENIQKELNQIKIEKSQK 596

QY 151 --EMIGLQERDRQLQCNRLHQLLKNKEDVOKLQNI-----TASRATQYNDV 198
Db 597 EELNKIKKEQKQVEDEKAKLITDIANGNDGLTKLNEVIDKLKDEKENISNELNQIKNER 656

QY 199 KRKREYNKLERLHQ-----LVNKKDKNIAMDVLNVGVR----- 234
Db 657 DNISNFPNKTKEIKQENETIQLNEEKSVLLNELNQIKKEEKQIEDEKAVIQOEKENEI 716
```

```
QY 235 -----ADKRGSWRTDKTEARN-----EDENVKIL-----LN 261
Db 717 TKLNEDKTVIENELNQIKTEKOEIENELNQTKDEKQKIEDEKSKLITELSNNGDISKLN 776
QY 262 DYEYRQKQILMENAEALKVLQMKKEMISLLSPQKKKPRERABDGTGTV-----AIS 313
Db 777 BELTQTKQ--EKENVNLNQLKNFASFQKQNTQKENELKDNKNKVOQELQKKNVEVS 833
QY 314 DIEDSGELSRDSVWGLSCDTVRE-----QLTNSIRKOWRILKSHVEKLDNQASKV 364
Db 834 KLEEEKGNISNE---LS-NTKQELQKQKEIITITQEKKEKENELKEQVKKIEEEKSKL 888
QY 365 HSE-----GLANEEDVISRODH-----EQETEKLELEIERCKEMIKAQOQLLOQ 408
Db 889 ITELSNNGSDGISKJNLBELTQTKOEKEIQKALEEKEKLEKETELKTEKAKQELKEEK 948
QY 409 LATTCCDDTTSLRDCYLLLEEKERLKEEWTLFKQEKKNFERRRSFTAAIRLGLERKAF 468
Db 949 NKTI--BEKTNLQOE---LNENKKIVEELTQTKOEKEEINNELNSIKE-----EKRI 996
QY 469 EEERASVWVKQQLNMTNFDHQNSENVK 495
Db 997 EEE-----KNQIINENK--EIKEENIK 1016
```

Search completed: June 12, 2006, 19:14:00  
Job time : 306 secs

**THIS PAGE LEFT BLANK**

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:15:24 ; Search time 182 Seconds  
(without alignments)  
1565.259 Million cell updates/sec

Title: US-10-644-084-2  
Perfect score: 3165  
Sequence: 1 MGDWMTVDPVLTCTENKNLS.....CYSGCSSAFSAHGDRDDL 615

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2152.5	68.0	504	4	US-10-104-047-3467
2	2152.5	68.0	504	6	US-11-072-512-3467
3	379.5	12.0	443	4	US-10-425-114-54149
4	379.5	12.0	378	4	US-10-424-599-170119
5	377	11.9	389	4	US-10-425-115-341817
6	290	9.2	290	4	US-10-425-114-45937
7	243.5	7.7	1985	5	US-10-732-923-3351
8	238	7.5	2135	6	US-11-203-806A-12
9	234.5	7.4	1679	4	US-10-369-493-22080
10	233	7.4	206	4	US-10-767-701-38884
11	233	7.4	987	4	US-10-437-963-180170
12	226.5	7.2	1175	5	US-10-732-923-3302
13	225	7.1	2503	5	US-10-828-985A-11
14	222.5	7.0	3225	4	US-10-408-765A-254
15	220.5	7.0	1875	4	US-10-369-493-23285
16	220.5	7.0	1875	5	US-10-732-923-3334
17	220.5	7.0	1875	5	US-10-732-923-3335
18	218.5	6.9	1128	5	US-10-732-923-3317
19	218	6.9	2383	4	US-10-082-830-260
20	217.5	6.9	689	4	US-10-108-605-305
21	217.5	6.9	744	6	US-11-097-143-40755
22	217.5	6.9	746	5	US-10-732-923-3339
23	217.5	6.9	1379	4	US-10-205-219-5
24	217.5	6.9	1909	5	US-10-732-923-3341
25	216.5	6.8	1132	5	US-10-732-923-3315
26	216.5	6.8	2665	6	US-11-124-368A-214
27	216.5	6.8	2668	6	US-11-124-368A-215

ALIGNMENTS

RESULT 1  
US-10-104-047-3467  
; Sequence 3467, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3467  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-104-047-3467

Query Match 68.0%; Score 2152.5; DB 4; Length 504;  
Best Local Similarity 90.3%; Pred. No. 3.6e-132;  
Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;  
QY 1 MGDWMTVDPVLTCTENKNLSQVTSKTPSSLYSQOVLCSVPISKNVHGVGFCTGE 60  
DB 1 MGDWMTVDPVLTCTENKNLSQVTSKTPSSLYSQOVLCSVPISKNVHGFCTGE 56  
QY 61 NIEQISYLDQELTTFGFPSPSYEESKSEAKRELINAVLNCNMLLVLRKNLLAQESV 120  
DB 57 NIEQISYLDQELTTFGFPSPSYEESKSEAKRELINAVLNCNMLLVLRKNLLAQENV 116  
QY 121 ETQNLKSGMDHLQSCYAKLKEQLETSRREMIGLOERDRLQCKNRSHLQLKNEKQSV 180  
DB 117 ETQNLKSGMDHLQSCYAKLKEQLETSRREMIGLOERDRLQCKNRSHLQLKNEKQSV 176  
QY 181 OKLQNIASRATQYNDHDKREYKNLKERHQLVMNKKDKNIAMDVLNVYGRADKRG 240  
DB 177 OKLQNIASRATQYNDHDKREYKNLKERHQLVMNKKDKNIAMDVLNVYGRADKRG 236  
QY 241 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 300  
DB 237 SWRTDKTEARNEDEMYKILLNDYEQKQILMENAELKKVLQOMKEMISLLSPQKKPR 296  
QY 301 ERAEDGTGTVASIDTDDSGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHEKLDNQ 360  
DB 297 ERVDDSTGTV-ISDVEEDAGELSRDVSVMGLSCDTVREQLTNSIRKQWIRLKSHEKLDNQ 355  
QY 361 ASKVHSEGLNEEDVLSRODHEQETKLEITERCKEMIKAAQOOLLOOQLATTCDDDTSL 420

```
Db 356 VSKVHLEGFNDENVISRODHEQTEKLELEIQCKEMIKTQOQLQQOQLATAYDDTTSL 415
QY 421 LRDCYLLEKERLKEEWTLFKQKKNFERRRSFTEAAIRLGLE 464
Db 416 LRDCYLLEKERLKEEWSLFKQKKNFERRRSFTEAAIRLGLE 459

RESULT 2
US-11-072-512-3467
; Sequence 3467, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHITO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3467
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3467

Query Match 68.0%; Score 2152.5; DB 6; Length 504;
Best Local Similarity 90.3%; Pred. No. 3.6e-132;
Matches 419; Conservative 20; Mismatches 20; Indels 5; Gaps 2;

QY 1 MGDWMTVTPVLCTENKNLSOYTSKSPSSLYSQOVLCSVPVLSKNVHGVGVCCTGE 60
Db 1 MGDWMTVTPD-----BSKTSIQYTSKSPSSLYSQOVLCSVPVLSKNVHGSFSAFCTED 56
QY 61 NIEQISYLDQLTTFGPPSLYEESKREKELNIIVAVLNCMNEILLVLRKNLLAQBSV 120
Db 57 NIEQISYLDQLTTFGPPSLYEESKREKELNIIVAVLNCMNEILLVLRKNLLAQENV 116
QY 121 ETQNLKSGMDHLOSCYAKLKEQLETSRREMIGLQERDRQCKNRSRLHQLLNKNEDEV 180
Db 117 ETQNLKPGSDMDHLOSCYKLEQLETSRREMIGLQERDRQCKNRSRLHQLLNKNEDEV 176
QY 181 OKLQNIIASRATQYNHDKRERENYKLERHQLVNMKDKNIAMDVLNYYGRADGKRG 240
Db 177 OKLQNIIASRATQYNHDKRERENYKLERHQLVNMKDKNIAMDVLNYYGRADGKRG 236
QY 241 SWRTDKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMCKEMISLLSPKKKPR 300
Db 237 SWRTCKTEARNEDEMYKILLNDYERQKQILMENAELKKVLOQMCKEMISLLSPKKKPR 296
QY 301 ERAEDGTGTVAISDIEDSGELSRDSVNGLSCTVREQLTNSIRKQWRILKSHVEKLDNQ 360
Db 297 ERVDDSTGTV-ISDVEEDAGELSRDSMDLSCTVREQLTNSIRKQWRILKSHVEKLDNQ 355
```

```
QY 361 ASKVHSEGLNEEDVTSRODHEQTEKLELEIERCKEMIKAOOQLQQOQLATTCCDDTTSL 420
Db 356 VSKVHLEGFNDENVISRODHEQTEKLELEIQCKEMIKTQOQLQQOQLATAYDDTTSL 415
QY 421 LRDCYLLEKERLKEEWTLFKQKKNFERRRSFTEAAIRLGLE 464
Db 416 LRDCYLLEKERLKEEWSLFKQKKNFERRRSFTEAAIRLGLE 459

RESULT 3
US-10-425-114-54149
; Sequence 54149, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54149
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-041-D8_FLI pep
US-10-425-114-54149

Query Match 12.0%; Score 379.5; DB 4; Length 443;
Best Local Similarity 26.6%; Pred. No. 2e-16;
Matches 106; Conservative 101; Mismatches 135; Indels 57; Gaps 14;

QY 27 KMPSSLY-----SQOVLCSVPVLSKNVHGVGVCCTGENIEQISYLDQLTTFGPPSL 81
Db 54 RMSSARFDLRASSQL--QPPPSHASMSDGGAFANAENLEHCARYLNQTLVTFGPPA- 110
QY 82 YEESSKREKELNIIVAVLNCMNEILLVLRKNLLAQBSVETQNLKLGSDMDHLOSCYAKL 141
Db 111 ---SLDLFATDPVSIARTCNCIYALLOQRDIEFRSTNDQRQRMQSDISRLEAKIERM 167
QY 142 KEQLETSRREMIGLQERDRQCKNRSRLHQLLNKNEDEVQKQNIIASR--ATQYNHDKV 199
Db 168 DAQLAAKDRELATLTRTEAKNTAALKSQIDKLOQERDEFQKM--VIGNQVVRTQIHEMK 225
QY 200 RKEREYNKLERLHQLVNMKDKN--TAMDVLNYYGRADGKRGSRWTDKTEARNEDEMYK 257
Db 226 KKEKEYIKLOEKLQVLMEKKESRSGMIMNLQKEGRQRTGWNGKK---NDNDYYK 281
QY 258 ILLNDYERQKQIILMENAELKKVLOQMCKEMISLLSPKKKPREAEDGTGTVAISDI-E 316
Db 282 MIVDAYEVKKQELMQENADLRLLRSQMOMRDFLN-----APNGSSQSTVTDNGR 332
QY 317 DSDGEL-----SRDSVNGLSCTVREQLTNSIRKQWRILKSHVBEKLDN--QASKVHSEGL 369
Db 333 QESSGPSPLGGKTDVFDLPFPMARDQIEESLRTKMTSIKARMTQLODAQKGAEVTSEAT 392
QY 370 NEEDVISRODHEQTEKLELEIERCKEMIKAOOQLQQO 408
Db 393 DRE-----LELE---AQLVEA-RSIIQEQ 412

RESULT 4
US-10-424-599-170119
; Sequence 170119, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
```



```

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 170119
; LENGTH: 378
; TYPE: prt
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_124632C.1.psp
US-10-424-599-170119

```

Query Match	12.0%;	Score 379;	DB 4;	Length 378;
Best Local Similarity	26.3%;	Pred. No. 1.8e-16;		
Matches	94;	Conservative 95;	Mismatches 143;	Indels 26; Gaps 8;
QY	56	FCGTENLEQSIYLDQELTTTFGPPSLYEKSKAEKRELNIIVLVLCMNBELLVLQKNLL	115	
DB	25	FADADNLEHCAKYLNQSLVTFGPPA----SLDUFANDPVSFARTCNCIYVLLQQRQDVE	80	
QY	116	AQBSVEQNLKLGSDMBHLOSCYAKLKEQLSETRSMIGLQERDROLQCKNRSLHOLLKN	175	
DB	81	FRESANDORQRLSLDISLEAKYVERLEGQVQKDREIATITRTFAKNTAALKKAQIEKLOQ	140	
QY	176	EKEDEVKQLNIIASR--ATQYNHDVKKREYKNLKERLHQLVM-NKKDKNIAMDVLNVY	232	
DB	141	ERDEFQRM--VIGNQVQKTOQMHEMKKEKEYIKQLERLQVLMKEKKESRSGWEIMNLL	198	
QY	233	GRADGKRGSWRTDKTEARNEDEMYKILLNDYEVQKOILMENAEKVKVLOOMKEMISLL	292	
DB	199	QKSGRQGTWNGKA----DNDFYKILVDAYESKNQELMAENADRLLRSMQVDMRDFL	254	
QY	293	SPOKKPRPRAEDGTGTVAISDIEDDSGELSRDSVWGLSDCTVREQLTNSIRKQWRILKS	352	
DB	255	NAPNGLPK---QSFTGNVERVESDPQSPLVGKMDVFDLPFHMAQDQEEISLRNKWASIKE	311	
QY	353	HVEKLDNQASKVHSEGLNEEDVISRDQHEQTEKLELEIERCKEMIKAAQOOLQQOLA	410	
DB	312	RMVQL-----QDAKEAFTVSEATESELE-LEAQLVEARSIIQEQASIMSKHIA	359	

```

RESULT 5
US-10-425-115-341817
; Sequence 341817, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341817
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74904C.1.psp
US-10-425-115-341817

```

Query Match	11.9%	Score 377;	DB 4;	Length 389;
Best Local Similarity	27.0%	Pred. NO. 2.5e-16;		
Matches 99; Conservative	95;	Mismatches 123;	Indels 50;	Gaps 12;

```
Qy   54 GVFCCTGENIEOSISVYLDQEILTFGPPSPSYEESKSKAEKRELNIVAVLNCMELLVLQRKN 113  
      |||    |||    |||    |         ||     ||       || :||  
Db   30 GAFANAENLEHCARYNLQTLVTFGPFA-----SLDLFATDPVSIARTCICIALLQQORQD 85  
      |||    |||    |||    |         ||     ||       || :||  
Qy  114 LLAQSVETQNKLGSMDHQLSCVAKYLKEOLETSRREMIGLOERDRQLCKNRSIHOLL 173  
      |||    |||    |||    |||    |||    |||    |||    |||  
Db   86 IEFRESTNDQRWQSISRUEAKERMQDAQAANKRELATLTRTEAKNTAAKSQIDKL 145  
      |||    |||    |||    |||    |||    |||    |||    |||  
Qy  174 KNEDEVOQLNIIASR--ATQYNHDVKRKEREYNKLERHLQHLYMNKKDKN--IAMDVL 229  
      |||    |||    |||    |||    |||    |||    |||    |||  
Db  146 QOEDEFQKM-VIGNQOVRTQTTHMKKKEYIKLQELKNVLMEKKSSRSSGMETM 203  
      |||    |||    |||    |||    |||    |||    |||    |||  
Qy  230 NYVGCRADKGSRWTDTKEARNEDEMUYIKLLNDYEYROKOILMENAEUKVUQQKKEMI 289  
      |||    |||    |||    |||    |||    |||    |||    |||  
Db  204 NLLQEGRGORTWGNGK---NNDNYYKMIVDAYEVKKQELMQENADLRALRSQMOMNR 259  
      |||    |||    |||    |||    |||    |||    |||    |||  
Qy  290 SLLSQPKKPRAEDGTGTVAISDI-EDDSGL-----SRDSVMGLSODTVREOLTNSI 343  
      |||    |||    |||    |||    |||    |||    |||    |||  
Db  260 DFLN-----APNGSSQSTVTDNGRQSGSQPSPLGGKTDDFDLPFHWARDQIESL 310  
      |||    |||    |||    |||    |||    |||    |||    |||  
Qy  344 RKOWRIILKSHVEKLDN-QASKVHSEGLENEEDVISRODHQETEKLELEIERCKEMIKAQ 401  
      |||    |||    |||    |||    |||    |||    |||    |||  
Db  311 RTKWTSIKARTMLQDAQKAQEAFTSEATDRE-----LELE----AQLVER- 351  
  
Qy   402 QQLLQQQ 408  
      |||    |||    |||  
Db  352 RSIIQEQ 358
```

```
RESULT 6  
US-10-425-114-45937  
; Sequence 45937, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 45937  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700894116_FLI.pep  
US-10-425-114-45937
```

	Query Match	9.2%	Score 290;	DB 4;	Length 290;
	Best Local Similarity	25.0%;	Pred. No. 8.2e-11;		
	Matches	72;	Conservative	82;	Mismatches 112;
				Indels	22;
					Gaps 7;
Qy	126	KLGSMDHLOS	CYAKLKEQLET	SRREMIGLQERDR	QLOCKNRSUHLQLLKNEKDEYQKLQN 185
Db	3	RLLSDSIRL	EAKVERLEQLQ	VDREIATITRTEAKN	TAALKAQIEKLOQERDEFQRM-- 60
Qy	186	IIASR--AT	QYNHDVKKEREY	NKLERLHOLVM--	NKKDKNIAMDVLNYVGRADKRGSW 242
Db	61	VIGNQVK	TOQHEMKKKE	YIKLQERLNOVL	MEKKKESRGSEIWNLLQEGRQRGTW 120
Qy	243	RTDKTEAR	NDEDEMYKILL	NDYVYRQKQIL	MENALKKVLOOMKEMISLSPQKKKPRR 302
Db	121	NGKKA----	DNDFYKXIV	DAYESKNQELMA	ENADLRALLRSQVDMRDFLNAPNGLPK-- 174
Qy	303	AEDGTGT	VAISDTE	DDSGELSRDS	VMWELSCDTVREQLTNSIRKQWRILKSHVEKLINOAS 362

Db 175 -QSFTGNERVEPSQSPVLGKMDVDFLPFHMDQIEBSLRNKVASIKERVMQL----- 228  
Qy 363 KVHSEGLNEEDVLSRODHEQETEKLEIERCKEMIKAAQQLLOQOLA 410  
Db 229 ----QDAQKAEVTSATERELE-LEAQLVEARSIIQEQASIMSKHLA 271

## RESULT 7

US-10-732-923-3351  
; Sequence 3351, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Egerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3351  
; LENGTH: 1985  
; TYPE: PRT  
; ORGANISM: Plasmodium yoelii yoelii  
US-10-732-923-3351

Query Match 7.7%; Score 243.5; DB 5; Length 1985;  
Best Local Similarity 19.3%; Pred. No. 1.1e-06;  
Matches 110; Conservative 118; Mismatches 180; Indels 161; Gaps 21;  
Qy 12 LCTENKNSLQYTSKMSPLSSYQOVLGSSVPLSKNVHGVFGVCTGTENIRQSISYL-D 70  
Db 453 VCOASKQMARLQHIIESRQMLETKML-----LKKK-----KENFEKVGMLAD 497  
Qy 71 QELTFGFPSPLYEE-----SKSKAKRELINIVAVLNCWELL 107  
Db 498 KEKEIDKIHQIKESKELSKINKNNETDEKEKYVNSIKSYDNAQKEL-----LDKQNECI 552  
Qy 108 VL-----ORKNLLAQESVETONLKLGSMDH 133  
Db 553 LIENCKSKLYEYDEKFGQFNKKIIEIEREKEIEOERKNIIEKKENMLNENRR-----EIDE 609  
Qy 134 LQSYAKLKEQLETSRREMIGLO-ERDQLOCKNRL-----HOLLKNEDEVQKLQNTIA 188  
Db 610 EKLMMNKEKNELEMLKKELESLEKEKKTIIDCEYNNLQNKBELARNERNLIK-ENELK 668  
Qy 189 SPATQYNHDVKKEREYNKLERHLQVNMKKD-KNIAMDVLNYYGRADGKKGSRWTDKT 247  
Db 669 NRIDKYN-----BLIDELNKNKKEIENDKMKMLNDI----- 699  
Qy 248 EARNDEMYKILNDYERQKQILMENAEELKKVLQOMKEMISLLSPKKPRERAED-- 305  
Db 700 ----QDERIK-JLNETNITKKE--NEKEINYKKEIIEKKERSIMNDVEKMKRLMEDIE 751  
Qy 306 GTGTVAISDIEDSGELSRDSVWGLSCDTVRQLTNSIRKQWRILKSHVFKLDNQASKVH 365  
Db 752 NTKNIMLEDMEXNTK-----IKBEIENDKMMIKNIEDEKEKYKTYLEESKF 798  
Qy 366 SGLNEEDVISRODHEQETEKLELSIERCKEMI-----KAAQQLLOQOLATTCDDTT 418  
Db 799 NENLEKESELOKKYDDENNRLOAEINNEKKINKERNLEKQKKVYDEDFNCKEYEE 858  
Qy 419 SLRDCYLLEKE-----RLKEEWTLFKQKKNF-----ERERSFTAAIRLGLERKA 467  
Db 859 DIRKYNMLEEENNNKRYMKQELENYKKNVLDIEEEKDKLYVQOEKINLEKENIL 918  
Qy 468 FFEERASWYKQFLNMTNFDHNSNVKL 496  
Db 919 VEKEQ---IDIELKFNKFNKFEKEENDIKI 944

## RESULT 8

US-11-203-806A-12

; Sequence 12, Application US/11203806A  
; Publication No. US20060051737A1  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Silverman, Joshua  
; TITLE OF INVENTION: Nucleic and Amino Acid Sequences of  
; FILE REFERENCE: Human Rf1 and Methods of Use Thereof  
; FILE REFERENCE: 600-1-298N  
; CURRENT APPLICATION NUMBER: US/11/203,806A  
; CURRENT FILING DATE: 2005-08-15  
; PRIOR APPLICATION NUMBER: 60/601,405  
; PRIOR FILING DATE: 2004-08-13  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 2135  
; TYPE: PRT  
; ORGANISM: Fugu rubripes  
US-11-203-806A-12

Query Match 7.5%; Score 238; DB 6; Length 2135;  
Best Local Similarity 21.2%; Pred. No. 2.7e-06;  
Matches 125; Conservative 122; Mismatches 228; Indels 116; Gaps 24;  
Qy 86 KSEAKRELINIVAVLNCMNEELLVLRKNLLAQESVETQNLKLGSDMDHLQSCYAKLKEQL 145  
Db 1522 QAQSAARKCGIAE--QKLNLDILSKNRQDILAE-----KLKEEFENAKKCAQEAQAK 1572  
Qy 146 ETSRREMIGLO-----ERDQLOCKNRLHOLLKNEKD-EVKLQNTIASRA 191  
Db 1573 EAEREAASLRQKAEAEKQKAAAEAAKQAKAEMLKKEAEVEVSKR---TAAQA 1629  
Qy 192 TQYNHD-----VKRKEREYNKLER-----LHOLVNMKKDKNTAMDVLN-YVGRA 235  
Db 1630 TALQKQKQADDEMAKQKQAEALQKSLVEKELTVVKLQLEKTEKO--MDVLDELQRV 1687  
Qy 236 DGRGSRWRTKTEARNEDEMYKILNDYERQKQILMENAEELKKVLQOMKEMISLLSPQ 295  
Db 1688 KGEVND-AIKQKQAEVEELSCKVTSEMLLKLKIEE-ENRHLLOKDKMKMOQLAE 1744  
Qy 296 KKKPRERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVRQLTNSIRKQWRILKSHVE 355  
Db 1745 AAKMKLLAEDA-----ARLSVEATELARQMAESDLAEQALSERKMLKEMQVSBAT 1798  
Qy 356 KLDNQASKVHSEGLNEEDVISR--QDHEQETEKLELSIERCKEMIKAAQQLLOQOLATTC 413  
Db 1799 KLKAEAELOQRNOAQEMANKLODKQKIQERLDQETEAQKSLAEQ---KRLQLEISA 1855  
Qy 414 DDDTTSL-----LRDCYLLEKE-----RLKEEWTLFKQKKNF-----E 448  
Db 1856 ETEALKLVKVELTDSNAKAESEVKKVQRSDVVKVQLQAEIKQNKIEVLQKSETOTLOSS 1915  
Qy 449 RERSFTAAIRLGLERKAFEEERASVVKQ-----QFLNMTNFDHNSNVKLFSAF 500  
Db 1916 REASLRKAVADLEKEREQLKKEAELOKQKSPSKLRDLLEALMGPPDVNHSPPSHVR--GTW 1973  
Qy 501 SGSDPDONLIV---HSRPRQKKLHSAVNGVPACTSK---LTKSLPASSTSDPFRTHSCV 554  
Db 1974 SPSASPSNLSLKSKQKPLEDEIPS-----PLVKSRRVSFADPQQOQETADDIRRSPCI 2028  
Qy 555 SEHSISVLNITTPESKPSSEVARESTDQKSVQSRPSREGCYS----GCSS 602  
Db 2029 RTSSPRKPRNASSSQPKISEMSQE-----PRPVSRDCVYPALVGCSA 2070

## RESULT 9

US-10-369-493-22080  
; Sequence 22080, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 22080  
LENGTH: 1679  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22080

Query Match 7.4%; Score 234.5; DB 4; Length 1679;  
Best Local Similarity 19.6%; Pred. No. 3.4e-06;  
Matches 145; Conservative 124; Mismatches 230; Indels 241; Gaps 28;

QY 15 ENKNLSQYT-----SETKSPSSLYSQVLCSSVPLSKNVHGVFGVCTGE--- 60  
Db 1010 ENKDLSELVIRLEKDAADCOAELTKTKSSLYSAQDL-----LDKHERKWEERKADYEREL 1064  
QY 61 --NIEOS-----ISYLDQELITFG-----PSPSYEESKSEAKRELNIVA 98  
Db 1065 ISNIEOTESLVRNSVLIEKVDDTAANNQKHLKLVSLFNSURHNSLETK----- 1117  
QY 99 VLNCMNELLVLQRNLLAESV-ETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIG--- 154  
Db 1118 LTTCKRELAFVKQKNDSELEKTINDLQRTQTLSEKEY--QCSAVIIDEFKDITKEVTQVMI 1175  
QY 155 -----LQERDQLOCKNRSUHLQKNEKDEVOQLNTIASRATQV---NHDVKRERE 204  
Db 1176 LKENNAIOLKSLKNVTEKREIYKQLNDROEISRLQRLDLOQKEQVINSNKILVYSE 1235  
QY 205 YNKLKERLHQLVMNKKKNIAMDVLNVGADKRGSRWTDKTEARNEDEMYKILLNDYE 264  
Db 1236 MECKORYQDLSQQQKND-----AQKDIKLTNE-----ISDLK 1269  
QY 265 YRQKQILMENAELKVLQMKKEMISLLSPQKKPRERAEDGTGTVAISDIEDSDGELSR 324  
Db 1270 GKLSAENANADLENKFNRLKKAHEKLDASKQ----- 1303  
QY 325 DSVWGLSCDVTREQLNRSIRKQWRILKSHVEKLDNOA-----SKVHSEGLNEED 373  
Db 1304 -----QAALTNELNE-----LKAIKDKLEQLDHFENAKVIDLDTKLKAHELQSED 1348  
QY 374 VISROPDEOET-----EKLLEIERCKE-MIKAAQQQL 404  
Db 1349 V--SRDHEKDTYRTLMEETIESLKKELQIFKTANSSSDAFEKLKVMNEKEDRIIDERTKE 1406  
QY 405 LQQLATTCDDDTTSLRLDCYLLEEXERLKEW-----TL--FKEQKNFERERRSPT 455  
Db 1407 FEKKQETLNKSTSS---EAEYSKDIETLKKEWLKEYEDETLLRIKAEENLKKRIRLPS 1463  
QY 456 EAIR--LGLERKAFEEERASWVKQ----- 479  
Db 1464 EERIQIISKREELBEERFKKLKENAGSLTFLDNKSGEDAEELWNPSKNSRPSA 1523  
QY 480 ---FLNMTVPDQNS-ENVKLFSAPSGS-----SDPDNLIVHSRPRQKLKLVANGVPAC 530  
Db 1524 VAGFINQKRLKPOEQLKYNKNDVFSNDQSMVTNKENNTVDSSAGNKAIPTFSGFKPFF 1583  
QY 531 TSKLTSLPASSTSDPROTHSCVSEHSGISVLNITPEE-----SKPSEVARESTDQ 582  
Db 1584 SSN-TSSLQOS--PQNFPTASQSNINTNAPRLTNIQPEAVKAAINFNSVTDLTNNSTDG 1640  
QY 583 KWSVQRSPSSRECGYSGCS 602  
Db 1641 AKITEIGSTSKRPIESGTS 1660

RESULT 10  
US-10-767-701-38884  
; Sequence 38884, Application US/10767701  
; Publication No. US20040172694A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10/767,701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 38884  
; LENGTH: 206  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C77766\_1.pep  
US-10-767-701-38884

Query Match 7.4%; Score 233; DB 4; Length 206;  
Best Local Similarity 31.5%; Pred. No. 2.8e-07;  
Matches 57; Conservative 43; Mismatches 71; Indels 10; Gaps 4;

QY 54 GVFCGTGENTEQSISYLDQELITFGPSPSYEESKSEAKRELNIVAVLNCMNELLVLQRKN 113  
Db 29 GAFANAGNLEHCARYLNQVLTVTFGPPA-----SLDLFATDPVSTARTCNCIYALQQQRQD 84  
QY 114 LLAQSVETQNLKLGSDMDHLQSCYAKLKEQLETSRREMIGLQERDQROCKNRSUHLQL 173  
Db 85 IEFRESTNDQRMQSDISRLEAKIERMDAQLAAKDRELATLTRTEAKNTAALKSQIDKL 144  
QY 174 KNEKDEVOQLNTIASR--ATQYNDHVKKEREYNKLERLHQLVMNKKDKN--IAMDVL 229  
Db 145 QERDEFFQK--VIGNQVQRTQOIHEMKKKEEYIKQBEKLNQVLMEKKKESRSGMEIM 202  
QY 230 N 230  
Db 203 N 203

RESULT 11  
US-10-437-963-180170  
; Sequence 180170, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 180170  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_77560C.1.pep  
US-10-437-963-180170

Query Match 7.4%; Score 233; DB 4; Length 987;  
Best Local Similarity 22.4%; Pred. No. 2.1e-06;  
Matches 151; Conservative 108; Mismatches 220; Indels 194; Gaps 35;



```
Db 462 DESVCSESVFNTLDLSIEWNPATKLNQ-----ENIESE 499
QY 66 ISYLDQELTFGPPSL---YEESKSEAKRELNIIVAVLNCMBLLVLRKNLAAQ-ESVE 121
Db 500 LNSLRAD-----YDNLVLDYEQURTEKEEMELKL-KEKNDLDEFEALERKTKKQDEESIE 553
QY 122 -----TONLKLGSMDHDLQSCYAKL-----KEQLET 147
Db 554 DPKOMQKTLFDAETVALDAKRESAFURSENLELEKMKELATTYKQWENDIQYQSLEA 613
QY 148 SRREMIGLOER-----DROIQCKNRSJHOLLKNEKDEVOK 182
Db 614 KKQWQVDLEKELQSAFNEITKLTSLDGKVPKDLLCNLELGKITDQLKELNKEVEENEA 673
QY 183 LQN--IIASRATQYNDHVKREREYNKLKERLHOLVNMKKO-----NIAMDVLNY 231
Db 674 LREEVILLSKSLPSEVERLRKEIQDKSEELH-IITSEKDKLFSVHHKESRVQGLLEE 732
QY 232 VGRADKRGSWRTDKTEARNEDEMYKILLNDYEROKQILMENAELK-----279
Db 733 IGKTKDDLATTQSNYKSTQDFONFKTLHMDFFQKYKMWLEENRNMQEIYNLSKBAQKF 792
QY 280 -----VLQOMKEMISLLSPKKKPRERA-----303
Db 793 DSSLGALKTSLVKTQLOEKTREVOERLNEHQLEQLENRDSPLQTVREKTLITEKL 852
QY 304 -----EDGTGTVAISDIE--DSGELSRSVWGLSCDTV-----REQLTNSIRKOWRIL 350
Db 853 QOTLEEVKTLTQEKDQLQLESQIERDQLKSDIHDVTVMNIDTQELRNALES---L 908
QY 351 KSHVEKLDNQASKVHSE-----GLNEE-----DV 374
Db 909 KOHOETINTLTKSISEVSNLHMEENTGETKDEFOQKMGVDKKQDLEAKNTQTLTADV 968
QY 375 ISRQHEQES-----TEKLEIEIERCKEM-----397
Db 969 KDNELIEQKIFSLIQEKNELOOMLESVIAEKEQLTKLKENIENTIENQELRLLGDE 1028
QY 398 IKAQOQLLOQ-----QIATTCDDTTSLRDYLLLEKERLKEEWTLFKE-----442
Db 1029 LKKQBEIVAQEKNAHAIKKEGELSRTCD-----RLAEVEKLEKESQOQLQEQOQL 1078
QY 443 -----QKK-----NFERERSFTEAAILRLGLERKAFEEERA 473
Db 1079 LNVQEMSEMQRKINSIENLKNELKNKELTLEHMETERLELAQKLMENYEEVKSITKERK 1138
QY 474 SSVKQOFLNWTNFDH-----QNSENVKLFSAFSGSSDDPDNLIHVSRRPKKL 520
Db 1139 VLKELQKSFETERDHLRGVIREIATGLQTKELKIAH-----IHLKEHQSTI 1186
QY 521 HSVANGVPACTSKLTKSLPASPTSDFRQTHSCVSEHSSISVLN-----ITPEESKPS 574
Db 1187 DELRRSVSEKTAQII-----NTQDLESHTKQLQE--EIPVLHBEQELLPNVKVSE 1235
```

## RESULT 14

US-10-408-765A-254

; Sequence 254, Application US/10408765A

; Publication No. US20040101874A1

; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.

; APPLICANT: Fahy, Eoin D.

; APPLICANT: Zhang, Bing

; APPLICANT: Gibson, Bradford W.

; APPLICANT: Taylor, Steven W.

; APPLICANT: Glenn, Gary M.

; APPLICANT: Warnock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408, 765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: PASCSEQ for Windows Version 4.0

; SEQ ID NO 254

; LENGTH: 3225

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-408-765A-254

Query Match 7.0%; Score 222.5; DB 4; Length 3225;  
Best Local Similarity 18.9%; Pred. No. 4.8e-05;  
Matches 139; Conservative 147; Mismatches 217; Indels 233; Gaps 32;

```
QY 15 ENKNLSQYTSYTKMS-----PSSLYSQVLCSSVP-LSKNVHGCVGFTGNIQ 64
Db 1191 ENENIGDQURQIQVRESIDGKLPT--DQOESCSSTPGLEEPFKATEQHHTQPVLES 1248
QY 65 SI-----SYLDQELTFFGPPSLVE---ESKSEAKRELNIIVAVLNCMBLLVLRKNL 116
Db 1249 NLCPDWPSPHSDASALOGGTSVAQIKAQKLEIAEKVELELKVSTTSSELTKKSEVFQL 1308
QY 117 QBSVETQNLKLS--DMHLSQCYAK-LKEQLETSRRMIGLOERDRLQCKNRSJHOLL 173
Db 1309 QEQINKQGLEIESLKTVSHEAEVHAESLQCKLESSQLIAGL-EHLRELQPK-----1359
QY 174 KNEKDEVOKLQNIASRATQYNDHVKREREYNKLKERLHOLVNMKKDNKNIAMDVLNYV 233
Db 1360 ---LDELOKL-----ISKKEEDSVYLSGQL-----SEKAAATKI---1391
QY 234 RADGKRGSWRTDKTEARNEDEMYKIL-----LNDYEROKQILMENAELK-----278
Db 1392 -----QTEIEQEDLIKALHOLENQAKHEHDERIKOLQVELCEMKQKPEIGE 1439
QY 279 --KVLQOMKEMISLLSPQK-----KKPRERAEDGTGV-----AISDIED--DSGELS 323
Db 1440 ESRAKQOIQRKLAALISRKALKENKSLQELSARGTIERLTSLADVESQVSAQNK 1499
QY 324 RDSVNG-----LSCDTVR-----EQL 339
Db 1500 KDTVLGRLLALQERDKLITEMDRSLLENQSLSSCESLKLALGLTEDKELVKEIESL 1559
QY 340 TNS-----IRKOWRIKLSHVEKLDNOASKVHS--EGLNES-----DVISR 377
Db 1560 KSKTAESTEWQEKHEKQKEYEILLQSYENVSNAEARIQHVVEAVRQEKQELYGKLRST 1619
QY 378 ODHEQETEK---LELEIERCKEMIK---AQOQLLOQO-----LATTTC 413
Db 1620 EANKKETEKQLQEAQOEMEEMKMKRPAKSKQKILELEEBEENDRLRAEVHPAGDTAKEC 1679
QY 414 DDDTTSLRDYLLLEKERLKEEWTLFKQEKONKFERERSFTEAAILRLGLERKAFEEERA 473
Db 1680 ME--TLLSSNASKMELEERVKMEYETLSKKFQSLMSEKDSLSEVQDLKHQIEGNVSKQA 1737
QY 474 SW-VKQOFLNMTNFDHNSENVKLFSAFSGSSDDPDNLIHVSRRPKKLHVSANGVPACTS 532
Db 1738 NLEATEKHQDNQTNVTEGTQSIIP-----GETEEQSLSMSTR-----PTCSE 1779
QY 533 KLTKSLPASPTSDFRQTHSCVSEHSSIS-----VLNITPEESKPSVARES 579
Db 1780 SVPSAKSANPAVSKDF-----SSHDEINNYLQIQIDQLKERIAGLKEEBKQKNKFSQTL 1832
QY 580 TDQKWSVQSRSSREG 595
Db 1833 ENKNTLLSQISTKDG 1848
```

## RESULT 15

US-10-363-493-22285

; Sequence 22285, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 22285  
LENGTH: 1875  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-22285

Query Match 7.0%; Score 220.5; DB 4; Length 1875;  
Best Local Similarity 20.2%; Pred. No. 3.2e-05;  
Matches 134; Conservative 119; Mismatches 238; Indels 171; Gaps 27;  
QY 16 NKNLSQYTSKMSPLSYQ-QVLCSS---VPLSKNVHGVGVFCT----- 58  
Db 1107 NSRIEDLSQNKL-----LYDQIQYTAADKEVNNSTNGPLNNIILITRRERDILDQVT 1162  
QY 59 -----GENIEQISYLDQELTTFGPSPLYEESK-----EAKRELNIVA-----VLNQM 103  
Db 1163 VAERDAKMLRQKISLMDVEL-----QDARTKLDNSRVEKENHSSIIQOHHDDIMEKL 1213  
QY 104 NELLVLQKXLLAQSVEQNLK---LGSMDHLOSCYAKLKEQLETSRREMIGLQERDR 160  
Db 1214 NQNLRLRESNITRLNELENNNNKKELQSELQKQNVAPIESLTKYSM---QEKEQ 1270  
QY 161 QLOCKNRSLOHLLKDEKQVOKLONLIASRATQYNHDKKEREYKLERLHQLVMNKK 220  
Db 1271 EKLAKVEVHRKKSQDILEHEQLSS-----DYEKLESELENKEELE-----NKE 1319  
QY 221 DKN1-AMDVNTYVGRADGRGSRWTDKTEARNEDEMYKILNDYEQKQILMENAEKK 279  
Db 1320 RGAEEAEKFNLRQAQER--LKTSKLSQDSLTQVNSLRDAKNVLENSLSEANARIEE 1377  
QY 280 VLOQMKKEMISLLSPQKKPRERAEDGTGTVAISDIEDSGELSRDSVWGLSCDTVR-EQ 338  
Db 1378 -LQNAKV-----AQGNQLEAIRKLQEDAERAKSRELOAKLEESTTSYES 1420  
QY 339 LTNSIRKOWRIKSHVEKLDNQASKVHSEGLNEEDVISRQDHEQETEKLELEIERCKEMI 398  
Db 1421 TINGLNEEITTLKEIEIKORIQOQLOQATSANEQNDLS---NIVSMKKSFEEDKIK-FI 1476  
QY 399 KAOQQLQOQLATTCDDDTTSLRDCYLLKEERLKEEWTL-FKEQKKNFERER----- 451  
Db 1477 KEKTQEVNEK-----ILEAQERLNQPSNINMBEIKKKWESEHEQEVSQ 1519  
QY 452 -----RSFTEAAILGLERKAFEEERASVWVKQPLAMTNFDHQNSENVKLF 498  
Db 1520 KIRAEAEALKKGIPLTEKINKIITERKKEELEK-----EFEEKVEERIKSME 1567  
QY 499 AFGSSD---PDNLIVHSRPROKHLHSVANGVPACTSKLTKSLPASPSTS----- 545  
Db 1568 Q-SGEIDVLRQLKQAKQKQKQKLENEYN-----KKLOEELKDVPHSSHISDDERDKL 1620  
QY 546 -----DPROTHSCVSEHS-----SISVLNITPEESKPFSEVARESTQKWSVQSRP 590  
Db 1621 RAEIESRLREEPNNELQAIKKKFDGKQQAAMKTTLLERKLAKMESQLSETKQSAESPP 1680  
QY 591 SS 592  
Db 1681 KS 1682

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 12, 2006, 19:16:24 ; Search time 9 Seconds  
(without alignments)  
863.952 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWTVDPVLCTENKLS.....CYSGSSAPRSHGRDDLP 615

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 64916 seqs, 12643201 residues

Total number of hits satisfying chosen parameters: 64916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	343	10.8	382	US-10-953-349-6160	Sequence 6160, Ap
2	258	8.2	284	US-10-953-349-6161	Sequence 6161, Ap
3	213	6.7	232	US-10-953-349-6162	Sequence 6162, Ap
4	206.5	6.5	3113	US-10-505-928-325	Sequence 325, Ap
5	205	6.5	972	US-10-480-962-4	Sequence 4, Appli
6	201	6.4	621	US-11-293-697-4409	Sequence 4409, Ap
7	188	5.9	498	US-10-953-349-20047	Sequence 20047, A
8	188	5.9	508	US-10-953-349-20046	Sequence 20046, A
9	188	5.9	544	US-10-953-349-20045	Sequence 20045, A
10	187	5.9	824	US-11-293-697-4573	Sequence 4573, Ap
11	185	5.8	2871	US-10-505-928-100	Sequence 100, App
12	181	5.7	561	US-10-953-349-20041	Sequence 20041, A
13	178.5	5.6	486	US-10-953-349-24039	Sequence 24039, A
14	177	5.6	1234	US-11-327-900-6	Sequence 6, Appli
15	176	5.6	950	US-11-293-697-4492	Sequence 4492, Ap
16	173.5	5.5	541	US-10-953-349-17059	Sequence 17059, A
17	173	5.5	512	US-11-293-697-4080	Sequence 4080, Ap
18	171	5.4	450	US-10-953-349-24040	Sequence 24040, A
19	169.5	5.4	440	US-10-953-349-24041	Sequence 24041, A
20	169	5.3	519	US-11-293-697-4747	Sequence 4747, Ap
21	168	5.3	944	US-11-293-697-3086	Sequence 3086, Ap
22	165.5	5.2	515	US-10-953-349-17060	Sequence 17060, A
23	164.5	5.2	531	US-10-953-349-20042	Sequence 20042, A
24	164	5.2	513	US-10-953-349-11130	Sequence 11130, A
25	164	5.2	602	US-10-953-349-11129	Sequence 11129, A

ALIGNMENTS

RESULT 1

US-10-953-349-6160  
; Sequence 6160, Application US/10953349  
; Publication No. US20060107345A1

GENERAL INFORMATION:

- ; APPLICANT: ALEXANDROV, Nikolai et al.
- ; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
- ; FILE OF INVENTION: ENCODED THERBY
- ; FILE REFERENCE: 2750-1579PUS2
- ; CURRENT APPLICATION NUMBER: US/10/953,349
- ; CURRENT FILING DATE: 2004-09-30
- ; NUMBER OF SEQ ID NOS: 40252
- ; SOFTWARE: Patent in version 3.3
- ; SEQ ID NO 6160
- ; LENGTH: 382
- ; TYPE: PRT
- ; ORGANISM: Arabidopsis thaliana

US-10-953-349-6160

Query Match 10.8%; Score 343; DB 6; Length 382;

Best Local Similarity 25.8%; Pred. No. 1.1e-10;

Matches 95; Conservative 92; Mismatches 139; Indels 42; Gaps 10;

QY	56	FTGTENISQSYLDQELTTGFPSPSYEESKSKAKRELINIVAVLNCNMELLVLQKRNLL	115
DB	28	FADVNLNLEKIKYLNQSLVTSQFSA----SLDLFATDPVSIARTCNVYALIQQRORDE	83
QY	116	AQESVETQNLKLGSDMDHLOSCYAKLKEQLETSSREMIGLOERDLOCKNLSLHLLKN	175
DB	84	FRESANDQRULLSDMARLEAKVERLETQLOKEREGLSVTTEAKNTAALTKQEKJQK	143
QY	176	EKDEVQKLQNTIAS--ATQYNHDVKRERENKRLKRLHQLVM--NKDKNTIAMDVNYV	232
DB	144	ERDEFQRM--VIANQVQTKQLHETKKEKEYIKQLERLQNLVLMKKETRSQMEIMNLL	201
QY	233	GRADKGRSWRTDKTEARNEDEMYKILLNDVEYRQKQILMENAELKVLQOMKEMISLL	292
DB	202	QKEGRQRTWSGKTKTDS----DFYKKIIVDAEAKNQELMAENTDLRALLRSTQGMRSFL	257
QY	293	SPQKKKPREREDG-----TGTVAISLDDSGELSRDSVWGLSCDTVREOLTNSJR	344
DB	258	N-----ASGGUTNOSLVANGRHGADPSQPLG--GKTDVFDLPFRMARGQLSDSJR	306
QY	345	KQWRIKSHVKEKLDNQAKVHSEGLNEEDVISRQHEQETBKELEIEIRCKEMIKAAQQL	404
DB	307	SKWYSIKERMGLVDQAQKEVS-----ITSEASERELE-LEAQLVEARSIIQOESI	356
QY	405	LOQLATT	412





Db 984 SLNQKNNLIQKSEFANYIDEREKSTSELSDOYKQEKLL-LQRCETGTNAVYEDLSQY 1042  
QY 408 QLATTCCDDTTSLRCYLLLEERLKEEWTLFKEQKKNFERRRSFTEAARLGLERKA 467  
Db 1043 KAAQKNNKLECLNCTSLCENRQNELE-----QLKEAFKHEQBF-----LTKLA 1089  
QY 468 FEERASWVK-----QQFLNMTNFQHNSENKVLFSAFSGSDPDNLIVHSRPROKKLH 521  
Db 1090 FAERNQNLMLLETVQALRSEMTQNNK-----SEAGLQKEIMTLKEEQNKWO 1142  
QY 522 SVANGVPACTSKLTKSLPASPTSDPROTHSCVSEHSSISVLNITPEESKPSSEVARESTD 581  
Db 1143 KEVNDLLQNEQLMKVM-----KTKHEC-----QNLSESPIN--- 1175  
QY 582 QKWSQSRPSSRECC 596  
Db 1176 ---SVKERESERNQC 1187

## RESULT 5

US-10-480-962-4  
; Sequence 4, Application US/10480962  
; Publication No. US20060115813A1  
; GENERAL INFORMATION:  
; APPLICANT: TANG, Y. Tom; BAUGHN, Mariah R.;  
; APPLICANT: LEE, Ernestine A.; RAMKUNAR, Jayalaxmi;  
; APPLICANT: YUE, Henry; GRIFFIN, Jennifer A.;  
; APPLICANT: CHAMLA, Narinder K.; TRAN, Bao  
; APPLICANT: GANDHI, Ameena R.; KHAN, Farrah A.;  
; APPLICANT: SWARNAKER, Anita; GURURAJAN, Rajagopal;  
; APPLICANT: POLICKY, Jennifer L.; YAO, Monique G.;  
; APPLICANT: WARREN, Bridget A.; GRIETZEN, Kimberly J.;  
; APPLICANT: ELLIOTT, Vicki S.; LEE, Soo Yeun;  
; APPLICANT: SANJANWALA, Bharati; HONCHELL, Cynthia D.;  
; APPLICANT: FORSVTHE, Ian J.; GORVAD, Ann E.;  
; APPLICANT: RICHARDSON, Thomas W.; LEE, Sally;  
; APPLICANT: BARROSO, Ines

; TITLE OF INVENTION: INTRACELLULAR SIGNALING MOLECULES

; FILE REFERENCE: PF-1002 USN

; CURRENT APPLICATION NUMBER: US/10/480,962

; CURRENT FILING DATE: 2003-12-04

; PRIOR APPLICATION NUMBER: PCT/US02/17955

; PRIOR FILING DATE: 2002-06-06

; PRIOR APPLICATION NUMBER: US 60/297,010

; PRIOR FILING DATE: 2001-06-08

; PRIOR APPLICATION NUMBER: US 60/298,706

; PRIOR FILING DATE: 2001-06-15

; PRIOR APPLICATION NUMBER: US 60/299,998

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 60/300,377

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 60/301,871

; PRIOR FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 60/303,349

; PRIOR FILING DATE: 2001-07-05

; PRIOR APPLICATION NUMBER: US 60/303,403

; PRIOR FILING DATE: 2001-07-06

; PRIOR APPLICATION NUMBER: US 60/351,927

; PRIOR FILING DATE: 2002-01-25

; NUMBER OF SEQ ID NOS: 40

; SOFTWARE: PERL Program

; SEQ ID NO 4

; LENGTH: 972

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No: 3418524CD1

US-10-480-962-4

Query Match 6.5%; Score 205; DB 6; Length 972;

Best Local Similarity 19.8%; Pred. No. 0.0021;  
Matches 126; Conservative 111; Mismatches 245; Indels 154; Gaps 24;

QY 63 EOSISYLDQELTTFGPPSLYEESKKEAKRELNIIVAVLNCMELLVLQRKNLLAQSSET 122

Db 176 ESKLSSMNSIKTFWSPEL---KKERALKDEASKITTIKQYRVVQENQHWMTIOA 231

QY 123 --QNLKLGSDMHL-----QSCYAKLKE-----QLETSRREMIGLQERDRQL 162

Db 232 LQDELRIQDLNQLFOODSSSRGTGEPVCAELTEENFORLHABHERQAKELFLLRKLEEM 291

QY 163 OCKNRSLLHLLKNEDEVOKLQNIITAS-----RATQYNHD-----VK 199

Db 292 ELRIETQKOTLNARDESINKLLEMQSKGLSAKATEEDHERTRRLAEAEHVHLLSLLE 351

QY 200 RKEREYNKLERLHQLVMNKKDNIAVDLVNVRADGKRGSGWRTDKTKEARNDEMYK-- 257

Db 352 QKEKENSMLREEMHRRFENAPDSAKTKALQTVIEMKDSKISSMERGLRDLLEEIOMLKSN 411

QY 258 --ILLNDYEROKQILMENAE---LKKVLQOMKKEMISLLSPQKKPRERAEDGTGTVAI 312

Db 412 GALSTEEREEMKQMEVYRSHSKFMKNKVEQLKEE---LSSKEAQWELKKKAAGLQA- 466

QY 313 SDIEDDSGELSDSVWGLSCDVTVRQLTNSIRKOWRILKSHVEKL-----DNQASKVH 365

Db 467 -EIGQVKQELSROTETLLALQTKLETN-----QFSDSKOHIEVLKESITAKEQRAAIIQ 521

QY 366 SE-----GLNEEDVI---SRQDHEQETEK-----LELEIE 392

Db 522 TEVDALRLRLLEKETMLNKKTKQIQDMAEEKQTQAGEIHLKMDLVDKERVNVLOKKIE 581

QY 393 RCKEMIKAAQQLLQO-----QLATTCDDTTSLLRDCYILLEEKERLKEEWTLFKEQK 444

Db 582 NLQEQLRDKQKQMSLKERVKSLQADTTNTDTALTTLLEA--LAEKERTIER---LKEQR 636

QY 445 KNFERE-----RRSFTEAAIRLGLERKAFEEERASW--VKQOFLMNTNPDHONSENV 494

Db 637 DRDEREKQEEIDNYKKDLKDLKEKVSLLQGDUSEKESALLDLKEHASSLASSGLKKDSRL 696

QY 495 KLFSAFSGSSDDPDNLIVHSRPRQKKLHVSANGVVPACTSKLTKSLPASPTSDPROTHSCV 554

Db 697 KTELEALEQKKEECLKMS--QLKKAHEAA-----LEARASPEMSDRIQ----- 738

QY 555 SEHSSISVLNITPESKSPS-----EVARESTDQK 583

Db 739 --HLEREITRYKDESSKQAQAEVDRLLEILKEVENEK 772

## RESULT 6

US-11-293-697-4409  
; Sequence 4409, Application US/11293697  
; Publication No. US20060105376A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: Novel full length cDNA  
; FILE REFERENCE: H1-A0106  
; CURRENT APPLICATION NUMBER: US/11/293,697  
; CURRENT FILING DATE: 2005-12-05  
; PRIOR APPLICATION NUMBER: US/10/108,260  
; PRIOR FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 5458  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4409  
; LENGTH: 621  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-293-697-4409

Query Match 6.4%; Score 201; DB 7; Length 621;

Best Local Similarity 19.2%; Pred. No. 0.0019;

Matches 97; Conservative 115; Mismatches 182; Indels 110; Gaps 20;

QY 12 LCTENKNSQYTSYTKMSPSSLYSQVLCSSVPLSKNVHGVFGVCTGCTENIEQSISYLDQ 71





Qy	544	-----TSDPRQTHSCVSEHSSISVLNIITPEES-----KPEVARESTDQK	583
		: : : : :	
Db	722	HSLNEELDFTVSYEEITIECADQLAISHQIAHLEERNKHLEDLIRKPREKARKPRSK-	780
		: : : : :	
Qy	584	WSVQSPRESS	592
		: :	
Db	781	-SLENNHPKS	780
		: :	

```

RESULT 11
US-10-505-928-100
; Sequence 100, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27
; PRIOR APPLICATION NUMBER: US 60/363,019
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: PatentIn 3.2
; SEQ ID NO 100
; LENGTH: 2871
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-505-928-100

```

Query March	5.8%	Score 185;	DB 6;	Length 2871;
Best Local Similarity	21.8%;	Pred. No. 0.073;		
Matches 109;	Conservative 88;	Mismatches 182;	Indels 120;	Gaps 22;
Qy	74	TTFGPSSLYEBSKSEAKRELNIIVA-----VLNCMNELLVLQRKNLLAQSVE	TQN 124	
Db	1219	TTIKESIMQKEDDSKNURNQUDRUSRENRLKDBIVRLNDSILOATQRRRAE	ENA-LQQ 1277	
Qy	125	LKLGS-MDLHQSACYAKLKEQL-----ETSRRMIGLQERDROLQCKNRS	HOLLKNEKDE 179	
Db	1278	KACGSEIMQKKHLEIELKQVMQQRSDNARHKQSLSEAAKTIQDKNKE	IERLKABEQEE 1337	
Qy	180	VQK---LQNIIASRATQYNHDV---KXREBEYNKLERLHOLVMNKKD-----	221	
Db	1338	AKRWEYENELSKVNNYDEIISLKNQGFETINIKTTTHQLTMQKEEDT	SGYRAQIDN 1397	
Qy	222	-----KNIAMDLVNYGR-----ADGRGSRWTDK-----TEA 249		
Db	1398	LTRNRSLSBEIKLKNLTLOTTENLRVRVEDIOQKATGSEVSQRQQLLEV	ELRVQTQM 1457	
Qy	250	RNEDEM-YKILLNDYERQKOILMENAEKVLQOMKKEMISLLSPQKKXP	R-ERAE--- 304	
Db	1458	RTEESVRYQSGLDD--AAKTIQDKNKEIERLKQLIDKETNDRKCLDEN	ARLQRVQYDL 1514	
Qy	305	---DGTGTVAISDIEDDSGELSRVSWGLSGD-----TVRQ-----	LTNSIRKQWRI 349	
Db	1515	QKANSATETINKLVQEQELTR-----LRIDYERSVQERTVKDQDIT	RFQNSL-KELQL 1568	
Qy	350	LKSHVEKLDNQASKVHSEGLNEEDVISRQDHQETEKLELEIERCKEM	IKAQ----- 401	
Db	1369	QKQVBEELNLRKKTAS-----EUSCKR-----KKLEBELEGMRSL	KEQAIKITNLT 1616	
Qy	402	QQLLQQQLATTCDDDTTSLRDCY---LLEEKERLKEEWTLFKEQ---	KKNFRERRRSFT 455	
Db	1617	QOLEQASIVKRRSEDDLROQRDVLGDHLRHKQRTQEBELRLRSSE	VALRQLLQEQESVK 1676	
Qy	456	EAAIRLGLERKAFEEERAS 474		
Db	1677	QAHLRNEHFQKAIEDKGRS 1695		

RESULT 12  
US-10-953-349-20041  
; Sequence 20041, Application US/10953349

```

; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POL
; TITLE OF INVENTION: ENCODED THERBY
; FILE REFERENCE: 2750-1579PUS2
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20041
; LENGTH: 561
; TYPE: PRT
; ORGANISM: Glycine max
US-10-953-349-20041

      Query Match      5.7%; Score 181; DB 6; Length 561;
      Best Local Similarity 20.6%; Pred. No. 0.016;
      Matches 128; Conservative 103; Mismatches 231; Indels 158; Gaps 24

QY 14 TENKNLSQYTSSETKMPSSLYSQVLCSSVPLSKNVHGVFGVCTCGENIEQSIYLDQEL 733
DB 14 TNELNLK-TSETQVKDMLALQNLAS-----TKEELEEKISELETAR 56

QY 74 TTFGPPSLYEESKKEAKRELNIIVAVLNCMN-ELLVLQR---KNLLAESVETQNLKLS 129
DB 74 TTFGPPSLYEESKKEAKRELNIIVAVLNCMN-ELLVLQR---KNLLAESVETQNLKLS 129

QY 57 S-----KLOEKEKLE-----SIEAALKSQEAQFLTVQEBLTKFKTEKLEATMEDLTR 106
DB 57 S-----KLOEKEKLE-----SIEAALKSQEAQFLTVQEBLTKFKTEKLEATMEDLTR 106

QY 130 DMDHLQCYAKLKEQLTSRREMIGLQERDRLOQCKRPSHLLKNKXDEVQKLQNIAS 189
DB 130 DMDHLQCYAKLKEQLTSRREMIGLQERDRLOQCKRPSHLLKNKXDEVQKLQNIAS 189

QY 107 SSKKPEELCADLEEKILSGENFRTDSLLSQALSNNAELQKVKSLED-LHNESGAAAA 165
DB 107 SSKKPEELCADLEEKILSGENFRTDSLLSQALSNNAELQKVKSLED-LHNESGAAAA 165

QY 190 RATQYNHDVKKREYKCLKERLHQLVMNKDKNIAMDVLNVVGRADGKRGSRWTDKTEA 249
DB 190 RATQYNHDVKKREYKCLKERLHQLVMNKDKNIAMDVLNVVGRADGKRGSRWTDKTEA 249

QY 166 TATQRSLELEGIQTSTAAREAKSQLELTRFIAEQRN-----VEL 209
DB 166 TATQRSLELEGIQTSTAAREAKSQLELTRFIAEQRN-----VEL 209

QY 250 RNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPRERADGTGT 309
DB 250 RNEDEMYKILLNDYEQKQILMENAEKLVQOMKEMISLLSPQKKPRERADGTGT 309

QY 210 EQQLNLVQLKTSDAEREVAELSEQISNLNAKLEEAKEEK-SLLNSQLQBYTEK----- 261
DB 210 EQQLNLVQLKTSDAEREVAELSEQISNLNAKLEEAKEEK-SLLNSQLQBYTEK----- 261

QY 310 VAISDIEDDSGLSRDSVWGSLSCDTVREQLTNSIRKQWRILKSHVKEKLDNQASKVHSEGL 369
DB 310 VAISDIEDDSGLSRDSVWGSLSCDTVREQLTNSIRKQWRILKSHVKEKLDNQASKVHSEGL 369

QY 262 --VALLESDLNQSS-----LRSSQLEBELKN-----VNEKCAEHEDRASMHRSR 305
DB 262 --VALLESDLNQSS-----LRSSQLEBELKN-----VNEKCAEHEDRASMHRSR 305

QY 370 NEEDVISR-----OPHEOETKEKLEIERCKEMIKAAQQLLQQLQATTCD----- 414
DB 370 NEEDVISR-----OPHEOETKEKLEIERCKEMIKAAQQLLQQLQATTCD----- 414

QY 306 ELEDLIQSSHKLESDSKVSELELLEAEKYRI---QELEQQIISTLEEKRGASEGOAN 361
DB 306 ELEDLIQSSHKLESDSKVSELELLEAEKYRI---QELEQQIISTLEEKRGASEGOAN 361

QY 415 ---DDTTSLLRDCYLL-----EERKLRKEBWTLPKEQKNFRERRSFTE 456
DB 415 ---DDTTSLLRDCYLL-----EERKLRKEBWTLPKEQKNFRERRSFTE 456

QY 362 KYLDDVSNLTSELAIQARASTLETTLQAANERNGKELEDSLNATVTEKKNLE-----D 414
DB 362 KYLDDVSNLTSELAIQARASTLETTLQAANERNGKELEDSLNATVTEKKNLE-----D 414

QY 457 AAILGLGERKAFEEERASWVKQFLNMTNPFQHSNENVKLFSAFSGSSDPDNLIVHSRPR 516
DB 457 AAILGLGERKAFEEERASWVKQFLNMTNPFQHSNENVKLFSAFSGSSDPDNLIVHSRPR 516

QY 415 ASI--SLNEKLAERKENLEILRDLNLTQDKLQSTE-----SDLR 452
DB 415 ASI--SLNEKLAERKENLEILRDLNLTQDKLQSTE-----SDLR 452

QY 517 QKLLHSAVANGVPACTSKLTKSLPASPSI-----SDFROTHSCVSHHSISVL--NITPEE 569
DB 517 QKLLHSAVANGVPACTSKLTKSLPASPSI-----SDFROTHSCVSHHSISVL--NITPEE 569

QY 453 EAEURE-----SEIEKKAASEENLVVRGRDIET---AARHSELQLLHESLTRDS 500
DB 453 EAEURE-----SEIEKKAASEENLVVRGRDIET---AARHSELQLLHESLTRDS 500

QY 570 SKPSEVARESTDQKWS-VQS 588
DB 570 SKPSEVARESTDQKWS-VQS 588

QY 501 EQKFEALEKFNKDKSEVQS 520
DB 501 EQKFEALEKFNKDKSEVQS 520

RESULT 13
US-10-953-349-24039
; Sequence 24039, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POL
; TITLE OF INVENTION: ENCODED THERBY

```



; SEQ ID NO 4492  
; LENGTH: 950  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-293-697-4492

```
Query Match          5.6%; Score 176; DB 7; Length 950;
Best Local Similarity 19.4%; Pred. No. 0.054;
Matches 124; Conservative 105; Mismatches 231; Indels 178; Gaps 26;

QY 60 ENIEQISVLDQELTFGPPSPSYESKSKAEKRELNIVAVLNCMNELLV-----LQRKN 113
Db 65 EKAQQILRATEEF-----KQLEEAQKKKISEAGKDLLLYKQLSGRLQLVN 110
QY 114 LLAQESVETQNLKLGSDMHLQSCYAKLQELTSRRMIGLQERDRLOLCKNRSL---- 169
Db 111 KURQALD-----LQLQEKQKEIAGKQKEIKDLQIAIDSLDSKD 151
QY 170 --HQLLKNEK-----DEVQKLQNIIASRATQYNHVKRREYNNKLERL--HQLVMN 218
Db 152 PKHSHMAQKSGKEQQLDIMNKYQQLSRLDEILSRIAKETEEIKDLEEQLTEGQIAAN 211
QY 219 ----KUD-KNIAMDVLNYVG-----RADKKGSGWRTDK-----TEARNEDEMYKILL 260
Db 212 EALKKDLGVISGLQELGTIRKQATQACNECRKLDEKETILLQRLTEVEQERDQLEIVA 271
QY 261 NDYEYRQKI-----LMENAEKVKVLOQMKEM-----ISLLSPQKKKPRERAE 304
Db 272 MDAENMRKELAEALQEHVNASLQQTQGDLSAYAELEARNLNRDAEANQKBELE 331
QY 305 DGTGTVAISDIEDDS--GELSRDSVWGLSCDTVREQLTNSIRKQW----- 347
Db 332 KVT---RLTQLSALQAELEKE-----RQALKNALGKAQFSEKEQENSELHAK 378
QY 348 -----RILKSHVEKLDNQASKVHSEGLNEEDVISQDHEQETEKL----ELEIERCKE 396
Db 379 LKHLQDDNNLLKQLKDFQNLHNLHVVDGLVRPEEVAARVDELRRKILKLTGEMNIHSPSD 438
QY 397 MIKAQQQLLQQLATTCCDDTTSLARDCYLLLE----KERLKEWTLFKEQKKNERERR 452
Db 439 VLGKSLADLQKQFSEI-----LARKWERDEAOVRERKKQEMALQOEKLATQOEFR 491
QY 453 SPTAAI--RLGLERKAFEEERASWVKQFLNMTNFDHQNSENVKLFSAFSGSSD----- 505
Db 492 QACERALEARNWFDKQHE-----ARIQWENEIHYLQENLAKSMEEIQGLTDLQLOE 543
QY 506 ----PDNLIVSR--PROKKLHNVAN-----GVPACTSKLTKSLPASPSSTDSPROT---H 551
Db 544 ADEEKERILAQURELEKKKKLEDAKSOEQVGLDKELKKLKAIVATSKLATAEITIAKD 603
QY 552 SCVSEHSSISVLNITPEES-----KPSEVAREST 580
Db 604 QLSLHGTVMKINQRAEELQEAERFSRKAQAARDLT 641
```

Search completed: June 12, 2006, 19:19:29  
Job time : 11 secs

GenCore version 5.1.1.9  
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 13, 2006, 23:11:04 ; Search time 891 Seconds  
(without alignments)  
7218.742 Million cell updates/sec

Title: US-10-644-084-2

Perfect score: 3165

Sequence: 1 MGDWMTVPVLTCTENKLS.....CYSGSSAFSAHGRRDRLP 615

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 5244920 seqs, 3486124231 residues

Total number of hits satisfying chosen parameters: 10489840

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlp  
-Q=/abss/ABSSWEB\_spool/US10644084/runat\_12062006\_150410\_1340/app\_query.fasta\_1  
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abss03p  
-USER=US10644084 @CGN\_1\_1147 @runat\_12062006\_150410\_1340 -NCFU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=3 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_8.\*

```
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*
15: Geneseqn2006s.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	3165	100.0 2692 12	ADO26490 Murine af
2	2937	92.8 3195 12	ADO26492 Rat afadi
3	2725.5	86.1 5476 10	ACD19444 CDNA enco

4	2720	85.9	2272	10	ADB63343	Adb63343 Human CDN
5	1982.5	62.6	2716	4	AAH14625	Aah14625 Human CDN
6	1192	37.7	826	4	AAH06926	Aah06926 Human CDN
7	856	27.0	1503	4	AAH16996	Aah16996 Human CDN
8	854	27.0	668	13	ADO78551	Adq78551 Novel can
9	715	22.6	563	4	AAH08911	Aah08911 Human CDN
10	671.5	21.2	498	2	AAV86670	Aav86670 EST clone
11	650	20.5	403	5	AAF66702	Aaf66702 Novel hum
12	635.5	20.1	428	3	ACAC3006	Aac03006 Human sec
13	380.5	12.0	1691	13	ADX36698	Adx36698 Plant ful
14	359.5	11.4	1556	14	AEB67285	Aeb67285 Rice geno
15	353	11.2	1544	3	AAC43689	Aac43689 Zea maye
16	343	10.8	1149	3	AAC46459	Aac46459 Arabidops
17	334.5	10.6	1062	13	ADR64454	Adr64454 Cotton CD
18	328	10.4	513	12	ACH72399	Ach72399 Human gen
19	326.5	10.3	1496	3	AAC33585	Aac33585 Arabidops
20	317	10.0	1485	3	AAC49413	Aac49413 Arabidops
21	310	9.8	1044	3	AAC49412	Aac49412 Arabidops
22	308.5	9.7	833	4	AAH03373	Aah03373 Human CDN
23	290	9.2	1199	13	ADX27288	Adx27288 Plant ful
24	285	9.0	178	12	ACH86155	Ach86155 Human gen
25	234.5	7.4	5040	13	ADT47329	Adt47329 Bacterial
26	225.5	7.1	7509	14	ADM44256	Adm44256 Human cen
27	224	7.1	2138	14	AEA51203	Aea51203 Chicken R
28	224	7.1	2943	3	AAH70229	Aah70229 Plasmodiu
29	223.5	7.1	10300	6	ABK70274	Abk70274 Human lun
30	223.5	7.1	10300	13	ADZ44554	Adz44554 Human bre
31	223.5	7.1	10300	14	ADZ26156	Adz26156 Human gen
32	223.5	7.1	10300	14	AED18268	Aed18268 Fibrotic
33	222	7.0	6619	5	ABA20636	Abas20636 Human ner
34	222	7.0	6619	5	ABA20638	Abas20638 Human ner
35	222	7.0	6619	5	ABA20637	Abas20637 Human ner
36	220.5	7.0	5628	13	ADT47534	Adt47534 Bacterial
37	220.5	7.0	6284	5	ABX71393	Abx71393 Human cel
38	220.5	7.0	6284	13	ADX97475	Adx97475 Pancreat
39	220.5	7.0	6284	15	AEB84913	Aee84913 IMPDH mar
40	220.5	7.0	6386	4	AAI57837	Aai57837 Human poi
41	219.5	6.9	2602	4	ABL29009	Abi29009 Drosophil
42	218	6.9	4747	13	ADQ87627	Adq87627 Human tum
43	218	6.9	7962	3	AAZ58980	Aaz58980 Human cyt
44	218	6.9	8347	6	ABK13313	Abk13313 DNA enco
45	218	6.9	8347	12	ADQ85937	Adq85937 Human tum

## ALIGNMENTS

RESULT 1  
ADO26490  
ID ADO26490 standard; DNA; 2692 BP.  
XX  
AC ADO26490;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Murine afadin dilution domain binding protein (ADIP) DNA SeqID 1.  
XX  
KW murine; mouse; gene; ds; afadin dilution domain binding protein; ADIP;  
KW afadin; actinin; binding inhibitor; cardiant; heat disease;  
KW myocardial infarction; myocarditis.  
XX  
OS Mus musculus.  
XX  
FH Key Location/Qualifiers  
CDS 80..1927  
FT /\*tag= a  
FT /product= "ADIP protein"  
XX  
FN JP2004135658-A.  
XX  
PD 13-MAY-2004.  
XX  
PF 14-AUG-2003; 2003JUP-00293554.  
XX

```
PR 27-SEP-2002; 2002JP-00284263.
XX (EISA ) EISAI CO LTD.
XX WPI; 2004-404616/38.
XX P-PSDB; ADO26491.
XX
XX New polynucleotide encoding an afadin dilution domain binding protein
XX having avidity with afadin or actinin, useful for diagnosing heart
XX diseases e.g. myocardial infarction.
XX
XX Claim 1; SEQ ID NO 1; 37pp; Japanese.
XX
XX This invention relates to a novel isolated nucleic acid encoding an
XX afadin dilution domain binding protein (ADIP) that exhibits an avidity
XX with afadin/actinin. Specifically, it refers to screening assays to
XX identify compounds that modulate ADIP avidity and provides suitable
XX agonists, antagonists and antibodies thereof. The present invention
XX provides methods to identify afadin and actinin binding inhibitors
XX therapeutically as cardiants to diagnose and/ or treat heart disease such
XX as myocardial infarction or myocardiitis. This polynucleotide sequence is
XX the murine ADIP DNA sequence of the invention.
XX
SQ Sequence 2692 BP; 814 A; 525 C; 691 G; 662 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1 96e-243 Length: 2692
Score: 3165.00 Matches: 615
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 12 Gaps: 0

US-10-644-084-2 (1-615) x ADO26490 (1-2692)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCyThrGluAsnLysAsnLeuSer 20
DB 80 ATGGAGATTTGGATGACTGTGACAGATCCAGTTCCTGTGTACAGAAAACAAAAATCTCTCT 139

QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCys 40
DB 140 CAATATACCTCAGAAACAAAGATGTCCTCGTCCAGATTTGTATCTCCAGCAAGTTCTGTGC 199

QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
DB 200 TCTTCAGTACCTTTATCCAAAACGTCATGCTGTTTTTCGGTGTCTTCTGCACAGGAGAG 259

QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80
DB 260 AACATTGAACAAAGTATTTCTCTATCTTGTATCAGGAGCTGACACCTTCGGGTTTCTCTCC 319

QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100
DB 320 TTGTATGAAGAAATCCAAAGTAAAGAGCAAGAGAGAAATTAATATATAGTCGCTGTTCTG 379

QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120
DB 380 AACTGTATGAACGAGCTGCTGCTGCTTCAGCGGAAGAACCTGCTGGCCACGAGAGCGTG 439

QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140
DB 440 GAGACACAGAACTTGAAGCTGGCAGTGCATGACATGGACCACTTCGACAGCTGTACGCCAAA 499

QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160
DB 500 CTTAAGGAGAGCTTGAACGTCACAGCGGGAGATGATCGGGCTTTCAAGAGAGAGACAGG 559

QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180
DB 560 CAGCTGCAGTGCAAGAACAGGAGTTTGATCAGCTCTCTGAAGATGAGAAAGATGAGGTA 619

QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200
DB 1700 TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTCTGAAACAGATTCATC 1759

DB 620 CAAAAAATTACAAAATATATCATAGCCAGCGGGCTACTCAGTATAATCATGATGTGGAAGAGG 679
QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220
DB 680 AAGGAGCGTGAATATAATAAGCTAAAGAGCGCTTCATCATCAGCTCGTTTATGAACAGAGAG 739
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240
DB 740 GATAAAAACATAGCATCGATGGATGTTTTAAATTTATGTGGTTCGAGCTGATGGCAACAGGCG 799
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260
DB 800 TCATGGAGGAGCTGCAAAAACAGAACCCAGGAATGAAGATGAGATGTACAAAATTTCTGTGTG 859
QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280
DB 860 AATGATTATGATGATACCCCGCAGACAGATCTCTGATGGAGAACCGGAGCTGAAGAGGTC 919
QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300
DB 920 CTCCAGCAGATGAAGAAGGAGATGATCTCTCTCTCTCTCTCAGAGAAGAACGCCAGG 979
QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320
DB 980 GAAAGAGCAGAGGACCGCAGGCACTGTGTCTATCTCCGATATAGAAGATGACTCTGGG 1039
QY 321 GluLeuSerArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThr 340
DB 1040 GAACGTAGCAGACAGACAGCGTGTGGGGCTTTCTCTGTACACTGTGAGAGAGCAGCTGACA 1099
QY 341 AsnSerIleArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360
DB 1100 AACAGCATCAGGAACACAGTGGAGAAATTTGAAAGTTCATGTAGAAAAAATCGATAACCAA 1159
QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380
DB 1160 GCTTCGAAGGTACACTCAGAGGGCTTTAATGAGGAGACGTCATCTCAGCAACAGACCAT 1219
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
DB 1220 GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1279
QY 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
DB 1280 CAGCAGCAGCTCTTTACAGCAGCAGCTGCCACCACTGTCATGATGATGACACCACTCACTG 1339
QY 421 LeuArgAspCysTyrLeuLeuGluLysGluArgLeuLysGluTyrThrLeuPhe 440
DB 1340 TTGGGAGACTGTTACTTTGCTGGAAAGAAAGAAACGCCCTTAAAGAGAGTGGACCTTTT 1399
QY 441 LysGluGlnLysLysAsnPheGluArgGluArgSerPheThrGluAlaAlaIleArg 460
DB 1400 AAAGAGCAAAAAGAAATTTGAGAGAGAAAGGCGAAGCTTTACAGAAGCTGCCATTGCA 1459
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPhe 480
DB 1460 TTGGGTTTGGAGAAAAGGCGCTTTGAAAGAGAGCGAGCAGCTGGTAAAGCAGCAGTTT 1519
QY 481 LeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPhe 500
DB 1520 TTAACATGACGAATCTTGACCACTCAGAACTCAGAAAATGGAACCTTTTTCAGTGCCTTC 1579
QY 501 SerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeu 520
DB 1580 TCAGGAAGTTCTGATCCAGACATCTTATAGTCCACTCAGCGCCACCGCAAGAGAGCTA 1639
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
DB 1640 CACAGTGTGGCTAATGGGGTGCCAGCTTGACATCAAAACTGACTAAATCTCTCTCTGCC 1699
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIle 560
DB 1700 TCACCTTCTACTTCAGACTTTTCGCCAGACACATTCATGTGTCTGAAACAGATTCATC 1759
```



QY 561 SerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThr 580  
 DB 1760 AGTGCTGCTGAATATACTCTCTGAAGAAAGTAAACCAAGTGAGGTTCCAAAGAGAAACGCG 1819  
 QY 581 AspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGlyCys 600  
 DB 1820 GATCAGAAGTGGAGCGTGCAGTCGAGCGCCAGCTCGCGGGAGGGTGCTACAGCGGATGC 1879  
 QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
 DB 1880 TCTTCGCGCTTCAGGAGCGCTCACGGGACCGAGATGACTTACCT 1924

## RESULT 2

AD026492  
 ID AD026492 standard; DNA; 3195 BP.  
 XX AC AD026492;  
 XX 29-JUL-2004 (first entry)  
 XX DE Rat afadin dilution domain binding protein (ADIP) DNA SeqID 3.

XX rat; gene; ds; afadin dilution domain binding protein; ADIP; afadin;  
 KW actinin; binding inhibitor; cardiant; heat disease;  
 KW myocardial infarction; myocarditis.

XX OS Rattus norvegicus.

XX FH Key Location/Qualifiers  
 FT CDS 79..1920  
 FT /\*tag= a  
 FT /product= "ADIP protein"

XX PN JP2004135658-A.

XX PD 13-MAY-2004.

XX PF 14-AUG-2003; 2003JP-00293554.

XX PR 27-SEP-2002; 2002JP-00284263.

XX PA (EISA ) EISAI CO LTD.

XX WPI: 2004-404616/38.

XX DR P-PSDB; AD026493.

XX PT New polynucleotide encoding an afadin dilution domain binding protein  
 PT having avidity with afadin or actinin, useful for diagnosing heart  
 PT diseases e.g. myocardial infarction.

XX XX Claim 1; SEQ ID NO 3; 37pp; Japanese.

XX CC This invention relates to a novel isolated nucleic acid encoding an  
 CC afadin dilution domain binding protein (ADIP) that exhibits an avidity  
 CC with afadin/actinin. Specifically, it refers to screening assays to  
 CC identify compounds that modulate ADIP avidity and provides suitable  
 CC agonists, antagonists and antibodies thereof. The present invention  
 CC provides methods to identify afadin and actinin binding inhibitors  
 CC therapeutically as cardiants to diagnose and/ or treat heat disease such  
 CC as myocardial infarction or myocarditis. This polynucleotide sequence is  
 CC the rat ADIP DNA sequence of the invention.

XX SQ Sequence 3195 BP; 929 A; 648 C; 794 G; 823 T; 0 U; 1 Other;

## Alignment Scores:

Pred. No.: 4, 45e-225 Length: 3195  
 Score: 2337.00 Matches: 567  
 Percent Similarity: 95.1% Conservative: 18  
 Best Local Similarity: 92.2% Mismatches: 28  
 Query Match: 92.8% Indels: 2  
 DB: 12 Gaps: 1

US-10-644-084-2 (1-615) x AD026492 (1-3195)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
 DB 79 ATGGGAGATTGGATGACTGTTACAGATCCAGTCTGTGTACAGAAAAACAAAATCTCTCT 138  
 QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
 DB 139 CAATATACCTCAGAAACAAAGATGTCCTCGTCAAGTTTATATCTCGCAGCAAGTACTGTGC 198  
 QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
 DB 199 TCTGCAACACCTTTATCCAAAGATGTGCATGGTGTTCAGTGCCTTCTGCACAGAGAG 258  
 QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
 DB 259 AACATCGAACACAGATATTTCGTATCTTGATCAGGAACCTGACTACCTTCGGTTTCCCTTCC 318  
 QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
 DB 319 TTGTATGAAGAAATCCAAAGTAAAGGAGCGCAAGCAGAGATTAAGTATATAGTTCTCTTC 378  
 QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
 DB 379 AACTGCATGAATGAGCTGCTTGTGTTTCAGCGAAGAACCTCTCTGCCCCGAGNAACGGTG 438  
 QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
 DB 439 GAGACACAGAAATCTGAAGCTGGGCAGTGACATGGACCACCTGCAGAGCTGCTACGCCAAA 498  
 QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 DB 499 CTTAAGGAACAGTTGGAGGCCCTCCAGCGCAGAGATGATCAGCCTTCAGAGAGAGACAGA 558  
 QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
 DB 559 CAGCTACAGTGCAAAAACAGGAATTTGCATCAGCTCTCTGAAAAACGAGNAAGANGTGA 618  
 QY 181 GlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
 DB 619 CAAAAATTTACAAAATATCATAGCCAGTCGGGCTACTCAGTATAATCATGATGTGAAGAGA 678  
 QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
 DB 679 AAGGAGCGGGAGTACAATAAACTGAAGGAGCGCTCGCATCAGCTTGTATTGAACAAGAAG 738  
 QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
 DB 739 GATAAAAATATAGCCATGCACGCTTTTAATTTACGTGGCCGAGTGGATGGAAAGCGAGGC 798  
 QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
 DB 799 TCCTGGAGGACTGATAAAACAGAACCCAGGAATGAAGATGAATGTACAAAATTCCTGCTG 858  
 QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
 DB 859 AATGATTATGATACCGCCAGAGCAGATCTCTGCTGGAGAAATGCGGAGCTGAAGAAGGTC 918  
 QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
 DB 919 CTCCAGCAGATGAAGAAAGAGATGATCTCTCTCTCTCTCTCAANAAGAGAAACCCAGA 978  
 QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
 DB 979 GAAAGAGCAGAGGACAGCACAGGCACTGTTGTCATCTCCGATGTAGAAGACGACGCTGGG 1038  
 QY 321 GluLeuSerArgAspSerValTrpClyLeuSerCysAspThrValArgGluGlnLeuThr 340  
 DB 1039 GAGCTGACAGAGATGGTGTGTGGAGCCTTTTCCTGTGACACTGTCTCAGGAGCAGCTTACA 1098  
 QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
 DB 1099 AACGACATCAGGAGCAGTGGAGAAATTTCTGAANAAGCCATGTGGAAAAAATCTGTATACCCA 1158

```
QY 361 AlaserLysValHisSerGluGlyLeuAenGluGluAspValIleSerArgGlnAepHis 380
Db 1159 GCTTCARAGGTACATCTCAGAGGCTTCATGAAGAGGAGCTCATCTCAGCACAGACCAT 1218
QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400
Db 1219 GAGCAAGAGACTGAGAAACTGGAGCTGGAGATTGAGCGGTGTAAGAGATGATCAAGGCT 1278
QY 401 GlnGlnGlnLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420
Db 1279 CAGCAGCAGCTCTTACAGCAACAGCTGGCCACTGCGTGTGATGACGACACACCTCCTCATG 1338
QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440
Db 1339 TTGCGAGACTGTACTTCTGCTGAAGAAAGAACGCCCTTAAAGAGAGTGGTCCCTTTT 1398
QY 441 LysGluGlnLysLysAspPheGluArgGluArgArgSerPheThrGluAlaAlaIleArg 460
Db 1399 AAAGAGCAAAAAAGAAATTTTTCAGAGAGAAAGACGAAAGCTTTTACAGAAAGCTGCTATTCCG 1458
QY 461 LeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnGlnPhe 480
Db 1459 TTGGGGTTGGAGAGAAAGCGTTTTCAGGAAGAGCGAGCCAGCTGGTGAAGCAGCAGTTT 1518
QY 481 LeuAenMetThrAsnPheAspHisGlnAenSerGluAsnValLysLeuPheSerAlaPhe 500
Db 1519 TTAAACATGACGACCTTTTGCATCACCAGAACTCAGAAATGTGAACATTTTCAGTGCCTTT 1578
QY 501 SerGlySerSerAspProAspAenLeuIleValHisSerArgProArgGlnLysLysLeu 520
Db 1579 TCAGGAAGTTCTGTATCCAGACAATCTTATAGTCCACCCACGCCACGGCAAGAAAGACCA 1638
QY 521 HisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla 540
Db 1639 CACATGTGCTGAATGGGGTGCCAGCTTGCACATCCAAACTGGCTTAAGTCTCTCCGACC 1698
QY 541 SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIle 560
Db 1699 TCACCT-----TCAGACTTCTGCCCGCTCGCTCATGTGTCTGAGCAGACAGTCCCGTC 1752
QY 561 SerValLeuAenIleThrProGluSerLysProSerGluValAlaArgGluSerThr 580
Db 1753 AGTGGCGCTGACTGACTCTCTGAAGAAACCAACCGAATGAGTGTGGAGAGAAAGTACG 1812
QY 581 AspGlnLysTrpSerValClnSerArgProSerSerArgGluGlyCysTyrSerGlyCys 600
Db 1813 GACCAAGAGTGGAGCGTGTGTCCAGACCACCTCCCGGAGGGTGTCTACGGTGGATGC 1872
QY 601 SerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeuPro 615
Db 1873 TCCTCGGCTACACAAGCTCCACCGTGGAGACGAGATGACTTACCA 1917
RESULT 3
ID ACID19444
XX ACID19444 standard; cDNA; 5476 BP.
XX AC ACID19444;
XX AC ACID19444;
XX AC ACID19444;
DT 25-AUG-2003 (first entry)
XX AC ACID19444;
DE AC ACID19444;
KW Human; NOV; gene therapy; endocrine related disease; diabetes;
KW Metabolism-related disease; obesity; central nervous system disorder;
KW Alzheimer's disease; Parkinson's disease; epilepsy; multiple sclerosis;
KW schizophrenia; depression; autoimmune disorder; inflammatory disorder;
KW psoriasis; allergy; lupus erythematosus; asthma; cancer;
KW inflammatory bowel disease; rheumatoid arthritis; osteoarthritis;
KW colon cancer; lung cancer; liver cancer; breast cancer; ovarian cancer;
KW prostate cancer; brain cancer; melanoma; liver disease; liver cirrhosis;
KW lung disease; emphysema; obstructive pulmonary disease; haemophilia;
KW stroke; infection; age; ss.
```

```
XX Homo sapiens.
OS WO2003023002-A2.
PN 20-MAR-2003.
PD 09-SEP-2002; 2002WO-US028539.
PF 07-SEP-2001; 2001US-0318120P.
XX 07-SEP-2001; 2001US-0318130P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322636P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324980P.
PR 17-APR-2002; 2002US-0373212P.
PR 06-SEP-2002; 2002US-00236177.
XX (CURA-) CURAGEN CORP.
PA Spyttek KA, Patturajan M, Gorman L, Li L, Anderson DW, Zhong M;
XX Shirlach VL, Vernet CAM, Ellerman K, Berghs C, Rothenberg ME, Guo X;
PI Shimkels RA, Leach MD, Catterton E, Kekuda R, Ji W, Miller CE;
PI Rieger DK, Taupier RJ, Shenoy SG, Liu X, Padigaru M, Alsobrook JP;
PI Lepley DM, Edinger SR, Burgess CE;
XX WPI: 2003-313242/30.
DR P-FSDB; ABO14751.
XX New cytoplasmic, nuclear membrane bound or secreted polypeptides (NOVX)
PT and polynucleotides, useful in gene therapy, e.g. for treating or
PT preventing obesity, multiple sclerosis, allergy, cancers, hemophilia,
PT stroke or infections.
XX Claim 20; Page 351-353; 586pp; English.
PS The invention describes a new isolated polypeptide (NOVX). The NOVX
XX polypeptide, nucleic acid and antibody are useful as therapeutics,
XX particularly in the manufacture of a medicament for treating a syndrome
XX associated with a human disease, which includes a pathology associated
XX with NOVX polypeptide. The DNA encoding the protein is useful in gene
XX therapy for treating the disease or condition. In particular, the NOVX
XX polypeptide or polynucleotide is useful for treating endocrine/
XX metabolism-related diseases (e.g. obesity or diabetes), central nervous
XX system disorders (e.g. Alzheimer's disease, Parkinson's disease,
XX epilepsy, multiple sclerosis, schizophrenia or depression), autoimmune
XX and inflammatory disorders (e.g. psoriasis, allergy, lupus erythematosus,
XX asthma, inflammatory bowel disease, rheumatoid arthritis or
XX osteoarthritis), cancers (e.g. colon, lung, liver, breast, ovarian,
XX prostate or brain cancers, or melanoma), liver diseases (e.g. liver
XX cirrhosis), lung diseases (emphysema or obstructive pulmonary disease),
XX haemophilia, stroke, or infections (e.g. viral, bacterial or parasitic).
XX These are also useful in developing powerful assay system for functional
XX analysis of various human disorders, as well as in diagnostic
XX applications, and for monitoring the effects of drugs during clinical
XX trials. This sequence encodes a novel human NOV protein
SQ Sequence 5476 BP; 1806 A; 887 C; 1063 G; 1720 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. NO.: 7.73e-208 Length: 5476
Score: 2725.50 Matches: 527
Percent Similarity: 93.1% Conservative: 34
Best Local Similarity: 87.7% Mismatches: 39
Query Match: 86.1% Indels: 1
DB: 10 Gaps: 1
```

US-10-644-084-2 (1-615) x ACD19444 (1-5476)		Db	1144	GAAGAAGCTTGATACCAAGTTTCAAAGGTACACCTGGAAAGTTTAAATGATGAAGATGTA	1203
QY	15	GluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyr	375	IleSerArgGlnAspHisGlnGlnThrGluLysLeuGluLeuGluLeuGluArgCys	394
Db	127	GAAGCAAAACTATCTCTCAATATACCTCAGAAACAAGATGTCCTCCATCAAGTTTATAC	1204	ATCTCAGACACAGACCATGAACAAGAACTGAAAACTCGAGTTAGAAATTCAGCAGTGT	1263
QY	35	SerGlnGlnValLeuCysSerSerValProLeuSerLysAsnValHisGlyValPheGly	395	LysGluMetIleLysAlaGlnGlnLeuLeuGlnGlnLeuGlnGlnLeuAlaThrThrCysAsp	414
Db	187	TCACAGCAAGTGTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAGT	1264	AAAGAAATGATTTAAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGTACTGCATATGAT	1323
QY	55	ValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThr	415	AspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLys	434
Db	247	GCTTCTGCACAGAGATATATTGNACAGAGTATCTCATATCTTGATCAGGAATTGACT	1324	GATGATACCACTTCACTATTACGAGACTGTTATTTGTGGAAAGAAAGAGAGCTCTCAA	1383
QY	75	ThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeu	435	GluGluTyrThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArgSerPhe	454
Db	307	ACTTTTGGTTTTCCTTCATTATATGAAGATCCAAAGGTAAGAGACAAAGAGAGTTA	1384	GAAGAATGGTCCCTTTTAAAGAGCAGAAAAAGAAATTTTGAGAGGAGAGACGAAGCTT	1443
QY	95	AsnIleValAlaValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsnLeu	455	ThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSer	474
Db	367	AATATAGTAGCTACTAAATTTGTATGAATGAGCTGCTTTGTGCTTCAGCGGAAGACCTT	1444	ACAGAAGCGCTATTTCGCTGGGATTGGAGAGAAAGGCAATTTGAAGAAGAAAGAGCAGT	1503
QY	115	LeuAlaGlnGluSerValGluThrGlnAsnLysLeuGlySerAspMetAspHisLeu	475	TrpValLysGlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnVal	494
Db	427	CTAGCTCAGGAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTA	1504	TGGTTAAAGCAGCAGTTCCTAATATGACTACCTTTGACCACCAAGAACTCAGAAAAATGTG	1563
QY	135	GlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgGluMetIleGly	495	LysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHisSerArg	514
Db	487	CAGAGCTGCTACTCAAACTTAAGGAACAACCTGGAAACCTCCAGCAGGGAATGATTGGG	1564	AAACTTTTCAGTCCTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTCACCTCGAGG	1623
QY	155	LeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLys	515	ProArgGlnLysLysLeuHisSerValAlaHisGlyValProAlaCysThrSerLysLeu	534
Db	547	CTTCAGGAAAGAGACAGACAGTTACAATGTATAAGAACAGGAATTTGCATCAGCTACTAAG	1624	CAGCGCAAAAGAACCTCACAGTGTCTAATGGGTCTCCAGTTTGCATGTCTAAACTT	1683
QY	175	AsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyr	535	ThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThrHisSerCysVal	554
Db	607	AATGAAAGATGAGGTGCAAAATTTACAAATATCATTTGCAAGTCGAGCTACTCAGTAT	1684	ACTAATCTCTTCCTGCTTCACCTTCCAGACTTTTCCGACACAGCTTCTCGTCATA	1743
QY	195	AsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGln	555	SerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLysProSerGlu	574
Db	667	AATCATGATATGACAGAAAAGAGCGTGAATATATAAATCACTGAAGAACGCTCATCAAA	1744	TCTGAACATAGTTCATCAATCAATGACTGAATATACTGCTGAAGAAATTTAAACCAATCAG	1803
QY	215	LeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArg	575	ValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGlu	594
Db	727	CTTGTTATGAACAAGAAAAGATAAGAAAATAGCTATGTCAGCAATTTTGAATATTGTTCGGAGA	1804	GTTGGAGAGAACCGTACAAATCAAAAATGGAGTGTGGCGTCAAGACCTCGATCACAGGAA	1863
QY	235	AlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGlu	595	GlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAspArgAspAspLeu	614
Db	787	GCTGATGAAAAGAGGGCTCCTGGAGGACTGGTAAAACTGGAAGCCAGGAATGAAGATGAA	1864	GTTGCTATAGTGGATGCTCCTTGAGCTACACAAATTTCTCATGTAGAAAAAGATGACTTA	1923
QY	255	MetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsn	615	Pro 615	
Db	847	ATGTATAAAATCTCTCTCAATGATTATGAATATCGTCAGAAACAATCTTAATGGAAAT	1924	CCT 1926	
QY	275	AlaGluLeuLysLysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerPro	RESULT 4		
Db	907	CGAGAACTTAAGAGAGTTCTTCAACAAATGAAAAAGGAAATGATTTCTCTCTTCTCC	ADDB63343		
QY	295	GlnLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAsp	ID	ADB63343 standard; cDNA; 2272 BP.	
Db	967	CAAAAGAAACCTTAGAAGAGATGATGATAGTAGTACAGGAACGTGTT---ATTTCAGT	XX	ADB63343;	
QY	315	IleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThr	XX	04-DEC-2003 (first entry)	
Db	1024	GTTGAAGAAGATGCGGGGAACCTAAGCAGAGAGATGATGGGACCTTTCCTGTGAAACT	DE	Human cDNA encoding clone TESTI20071130.	
QY	335	ValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisVal	KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;	
Db	1084	GTGAGAGACGCTTTACAAACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAATCATGTA	KW	tissue regeneration; cell regeneration; membrane protein;	
QY	355	GluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluAspVal	KW	signal transduction-related protein; transcription-related protein;	
			OS	osteoporosis; neurological disease; cancer; tumour.	
			XX	Homo sapiens.	
			FX	Key	Location/Qualifiers
			FT	CDS	224..1738

FT /\*tag= a  
 XX /product= "Clone TESTI20071130 protein"

PN EPI308459-A2.

XX 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

PR 25-JAN-2002; 2002US-00350978.

XX (HELI-) HELIX RES INST.

PA (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;

DR WPI; 2003-450961/43.

DR P-PSDB; ADB65313.

XX New polynucleotides and polypeptides, useful for developing a diagnostic

PT marker or medicines for regulation of their expression and activity, or

PT as targets of gene therapy.

XX Claim 1; Page: 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected

CC from 1970 fully defined nucleotide sequences which encode novel

CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide

CC or its partial peptide, an antibody binding to the polypeptide or peptide

CC of the polynucleotide, immunologically assaying the polypeptide or

CC peptide of the polynucleotide by contacting the polypeptide or peptide

CC with the antibody of the encoded protein, and observing the binding

CC between the two, a transformant carrying the polynucleotide in an

CC expressible manner and an antisense polynucleotide. The oligonucleotide

CC is useful as a primer for synthesising the polynucleotide, or as a probe

CC for detecting the polynucleotide. The polynucleotides and encoded

CC proteins are useful as pharmaceutical agents and many disease-related

CC genes may be included in them, for developing a diagnostic marker or

CC medicines for regulation of their expression and activity, or as targets

CC of gene therapy. The genes are involved in tissue and/or cell

CC regeneration. Membrane proteins, signal transduction-related proteins,

CC transcription-related proteins, disease-related proteins and genes

CC encoding them can be used as indicators for diseases (e.g. osteoporosis,

CC neurological diseases, cancer, tumours. The cDNA may be used to regulate

CC the activity or expression of the encoded protein to treat diseases. The

CC sequence presented is a cDNA of the invention. Note: Some of the sequence

CC data for this patent is not represented in the printed specification, but

CC is based on sequence information supplied by the European Patent Office.

Db 332 TCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAGAT 391  
 Qy 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
 Db 392 AATATTGAACAGAGTATCTCATATCTTGTATCAGAAATGACTACTTTTGGTTTCCCTTCA 451  
 Qy 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
 Db 452 TTATATGAGAGTATCCAAAGGTAAAGAGACAAGAGAGAGTTAAATATATAGTACTGTACTA 511  
 Qy 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
 Db 512 AATTGTATGAATGAGTCTTGTCTTCAGCGGAAGAACCTTCTAGCTCAGGAAATGTG 571  
 Qy 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
 Db 572 GAGACACAGAATTTGAAGCCGGGAAGTATATGACCATCTACAGAGCTGTACTCAAAA 631  
 Qy 141 LeuLysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArg 160  
 Db 632 CTTAAGGACAACTGGAAACCTCCAGGAGGAATGATTGGCTTCAGGAAAGAGACAGA 691  
 Qy 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluVal 180  
 Db 692 CAGTTACAATGTAAGAACAGCAATTTGCATCAGCTACTTAAAGAAATGAGAAAGATGAGGTG 751  
 Qy 181 GlnLysLeuGlnAsnIleLeuAspArgAlaThrGlnTyrAsnHisAspValLysArg 200  
 Db 752 CAAAAATTTACAAAATATCTATTGCAAGTCGAGCTACTCAGTATATATCATGATGAAGAGA 811  
 Qy 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
 Db 812 AAAGAGGTGNATATATATAAATCTGAAGAGCTCTACATCAACTTGTATGACACAGAAA 871  
 Qy 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
 Db 872 GATAAGAAAAATAGCTATGGACATTTTGAATTTATGTCGGGAGAGCTGATGGAAGAGAGGC 931  
 Qy 241 SerTrpArgThrAspLysThrGluAlaArgGlnAspGluMetTyrLysIleLeuLeu 260  
 Db 932 TCCTGGAGGAGCTGTGTAACCTGGAACCCAGGAAATGAAGATGAATGTATATAAATCTCTTG 991  
 Qy 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysVal 280  
 Db 992 AATGATTATGAATATCGTCAGAAACAATCCTAATGGAANAATGCAGAACTTAAGAAGGTT 1051  
 Qy 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
 Db 1052 CTTCAACAAATGAAAAAGGAAATGATTTCTCTTCTTCTCCCAAAAGAAACCTTAGA 1111  
 Qy 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
 Db 1112 GAAAGAGTAGATGATAGTACAGGAAGCTGTT---ATTTCCGATGTTGAAGAGATGCCGGG 1168  
 Qy 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
 Db 1169 GAACCTAAGCAGAGAGATATGTCGGACCTTTCTCTGTGAAACTGTGAGAGAGCAGCTTACA 1228  
 Qy 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
 Db 1229 AACAGCATCAGAAAAACAGTGGAGAAATTTTGAAGAGTCATGTAGAAAAACCTTGATACCAA 1288  
 Qy 361 AlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHis 380  
 Db 1289 GTTTCAAGGTTACACCTGGAGAGGTTTTAATGATGAAGATGTAATCTCAGCAAGACCAT 1348  
 Qy 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysLysGluMetIleLysAla 400  
 Db 1349 GAACAAGAACTGAAAACTCAGCTTAGAAATTCAGCAGTGTAAAGAAATGATTAAAACT 1408  
 Qy 401 GlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420

SQ Sequence 2272 BP; 783 A; 398 C; 502 G; 589 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 6.87e-208 Length: 2272  
 Score: 2720.00 Matches: 536  
 Percent Similarity: 86.8% Conservative: 34  
 Best Local Similarity: 81.6% Mismatches: 40  
 Query Match: 85.9% Indels: 47  
 DB: 10 Gaps: 3

US-10-644-084-2 (1-615) x ADB63343 (1-2272)

Qy 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20

Db 224 ATGGAGAGATTGGATGACTGTTACAGATCCA-----GAAAGCAAAACTATCTCT 271

Qy 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40

Db 272 CAATATACCTCAGAAACAAGATGTCATCAAGTTTATCTACACAAAGTGTATGT 331

Qy 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60

```
Db 1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGTCATATGATGATGATACCATCTTCACTA 1468
Qy 421 LeuArgAspCysTyrLeuLeuGluGluLeuGluArgLeuLeuGluGluTyrThrLeuPhe 440
Db 1469 TTACGAGAGCTGTTATTTGTTGGAAGAAAGAAAGCACTCTCAAGAGAAGATGGTCCCTTTT 1528
Qy 441 LysGluGlnLysLysAsnPheGluArgGluGluArgSerPheThrGluAlaIleArg 460
Db 1529 AAGACAGCAGGAAAGAAATTTTGAAGGAGAGAGCAAGCTTTTACAGAAGCCGCTATTGCG 1588
Qy 461 LeuGlyLeuGlu----- 464
Db 1589 CTGGGATTGGAGATTGGCTTTCTCTCGCCAAAGTGTTTTACACAAGAAATCTCTTTTGGGT 1648
Qy 464 ----- 464
Db 1649 ATCCATCATTCATCCAGCGTCACTAGTGTGTGTCAGAAGACAGTGCACCAACCAAGACTAT 1708
Qy 465 -----ArgLysAlaPheGluGluArgAlaSerTrpValLysGlnG 479
Db 1709 ATTGCTGTTTAAAAAAGAAAGCAATTGAAGAAGAAAGAGCCAGTGTGTTAAGCAGCA 1768
Qy 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499
Db 1769 GTTCTTAATATGACTACCTTTGACCACCAAGAACTCAGAAATGTGAACCTTTTCAGTGC 1828
Qy 499 aPheSerGlySerSerAspPheAspAsnLeuLeuValHisSerArgProArgGlnLysLys 519
Db 1829 CTTCTCAGGAAGTCTGATTGGGACAATCTTATAGTCACACTCGAGGCAGCGCAAGAA 1888
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539
Db 1889 GCCTCACAGTGTCTAAATGGGTCTCACTGTCATGTCATTAACCTTAACTTAACTTCTCTCC 1948
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559
Db 1949 TGCTTCACTTCCACTTCAGACTTTTGCCAGACACGCTTCTGCATATCTGAACATAGTTC 2008
Qy 559 rLysSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579
Db 2009 AATCAATGTACTGAATATAACTGCTGAAGAAATTAACCAAACTCAGGTTGGAGGAGACG 2068
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGl 599
Db 2069 TCAAAATCAAAATGAGTGTGGCGTCAAGACTGGATCAGAGCTGGATCAGGAGGTTGCTATAGTG 2128
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615
Db 2129 ATGCTCTTGAGCTACACAAATCTCATGTAGAAAAGATGACTTACCT 2177
```

## RESULT 5

```
AAH14625
ID AAH14625 standard; cDNA; 2716 BP.
AC AAH14625;
XX
DT 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:12264.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
```

```
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
```

## Claim 8; SEQ ID NO 12264; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

Sequence 2716 BP; 935 A; 405 C; 524 G; 852 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1e-148	Length:	2716
Score:	1982.50	Matches:	383
Percent Similarity:	92.6%	Conservative:	27
Best Local Similarity:	86.5%	Mismatches:	32
Query Match:	62.6%	Indels:	1
DB:	4	Gaps:	1

US-10-644-084-2 (1-615) x AAH14625 (1-2716)

Qy	173	LeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAlaThr	192
Db	3	CTAAAGATGAGAAAGATGAGGTGCAGAAATTTACAAATATCATTTGCAAGTGCAGTACT	62
Qy	193	GlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeu	212
Db	63	CAGTATAATCATGATATGAGAGAAAGAGCGTGAATATAATAAATGAAGCAAGCTCTA	122
Qy	213	HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal	232
Db	123	CATCAACTTGTATGAAACAAGAAAGATAAGAAAATAGCTATGAGCAATTTTGAATTTATGTC	182
Qy	233	GlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgGlu	252
Db	183	GGGAGAGCTGATGGAAAAGAGGCTCTGGAGGACTGGTAAACCTGAAAGCCAGGAATGAA	242
Qy	253	AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet	272
Db	243	GATGAATGTATAAAATTTCTCTTGAATGATATGATATCGTCAGAACAAATCTTAATG	302
Qy	273	GluAsnAlaGluLeuLysValLeuGlnMetLysLysGluMetIleSerLeuLeu	292

```
Db 303 GAAATGCAAGCTTAAAGAGCTTCTCAACAAATGAAAGGAAATGATTTCTCTCTT 362
Qy 293 SerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312
Db 363 TCTCCCAAGAAAGAACTTAGAGAAAGAGTAGATAGTACAGGAACGTGTT---ATT 419
Qy 313 SerAspIleGluAspSerGlyGluLeuSerArgAspSerValTIPGlyLeuSerCys 332
Db 420 TCCGATGTTGAAGAGATGCCGGGAACTTAGCAGAGAGAGTATGTGGACCTTTCCTGT 479
Qy 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
Db 480 GAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTTGAAGT 539
Qy 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372
Db 540 CATGTAGAAAGCTTGTATAACCAAGTTTCAAAGGTACACCTGGAAAGGTTTAAATGATGA 599
Qy 373 AspValIleSerArgGlnAspHisGluGlnGluThrGluLysLeuGluLysGluIleGlu 392
Db 600 GATGTAACTCAGCAGACCATGACAGAAACTGAAAGAACTCGAGTTAGAAATTCAG 659
Qy 393 ArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThr 412
Db 660 CAGTGTAAAGAAATGATTAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGTACTGCA 719
Qy 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGluArg 432
Db 720 TATGATGATGATACACACTTCACTATTACGAGACTGTATTGTTGGAAGAAAGGAACGT 779
Qy 433 LeuLysGluGluTrpThrLeuPheLysGluGlnLysLysAsnPheGluArgGluArg 452
Db 780 CTCAAAGAGAGATGTGCTCCCTTTTAAAGAGCAGAAAGAAATTTTGAGAGGAGAGACA 839
Qy 453 SerPheThrGluAlaAlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArg 472
Db 840 AGCTTTACAGAACCGCTATTGCGCTGGGATTTGGAGAGAAATGGCATTTTGAAGAAAGA 899
Qy 473 AlaSerTrpValLysGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGlu 492
Db 900 GCCAGTGTGTTAAGCAGCAGGTTTCTAAATATGACTACTCTTTGACCACCCAGAACTCAGAA 959
Qy 493 AsnValLysLeuPheSerAlaPheSerGlySerSerAspProAspAsnLeuIleValHis 512
Db 960 AATGTGAAACTTTTCAGTGCCTTCTCAGGAAGTTCGTATGGGACAACTTATAGTGAC 1019
Qy 513 SerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532
Db 1020 TCGAGCGACCGCAAAAGAGCCTCACAGTGTGTCTAATGGGTCTCCAGTTTGCATGTCT 1079
Qy 533 LysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThrHisSer 552
Db 1080 AAACCTTACTAAATCTCTTCCTGCTTCACCTCCACCTTCAGACTTTTGGCAGCAGCTTCC 1139
Qy 553 CysValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSerLysPro 572
Db 1140 TGCATATCTGAACATAGTTTCAATCAATGACTGTAATATACTGCTGAGAGAAATTAACA 1199
Qy 573 SerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
Db 1200 AATCAGGTTGGAGGAGATGTACAAATCAAAATCGAGTGTGGCGCTCAAGACCTGGATCA 1259
Qy 593 ArgGluGlyCysTyrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
Db 1260 CAGGAAGGTTGCTATAGTGGATGCTCCTTGAGCTACAAATTTCTCATGTAGAAAAAGAT 1319
Qy 613 AspLeuPro 615
Db 1320 GACTTACCT 1328
```

RESULT 6  
AAH06926

```
AAH06926 standard; cDNA; 826 BP.
AC AAH06926;
DT 26-JUN-2001 (first entry)
DE Human cDNA clone (5'-primer) SEQ ID NO:3761.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
OS Homo sapiens.
PN EP1074617-A2.
PD 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
DR WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX Claim 1; SEQ ID NO 3761; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences, and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX Sequence 826 BP; 329 A; 120 C; 179 G; 195 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 4,48-86 Length: 826
Score: 1192.00 Matches: 240
Percent Similarity: 92.7% Conservative: 14
Best Local Similarity: 87.6% Mismatches: 18
Query Match: 37.7% Indels: 4
DB: Gaps: 1
US-10-644-084-2 (1-615) x AAH06926 (1-826)
```

```
QY 173 LeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArgAlaThr 192
Db 3 CTAAGAAGATGAGAAAGATGAGGTGCAGAAATATACAAAATATCATTCGAAGTCAGACTACT 62

QY 193 GlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeu 212
Db 63 CAGTATAATCATGATATGAAGAGAGAGAAAGCGTGAATATAATAAACTGAAGGACGCTCTA 122

QY 213 HisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232
Db 123 CATCAACTTGTATGAACAAGAAAGATGAAGAAATAGCTATGACATTTTGAATTATGTC 182

QY 233 GlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGlu 252
Db 183 GCGAGAGCTGATGGAAAAGAGCGCTCTCGAGAGCTGGTAAACCTCAAGCCAGGAATGAA 242

QY 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272
Db 243 GATGAATCGTATAAAATCTCTTGAATGATTATGAATATCGTCAGAAACAATCTCTAATG 302

QY 273 GluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetLysSerLeuLeu 292
Db 303 GAAATATGCAAGACTTAAGAGAGGTCTTCAACAATGAAAAGGAAATGATTCTCTCTT 362

QY 293 SerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312
Db 363 TCTCCCAAGAGAAACCTAGAGAAAGAGTAGTAGTACAGGAACTGTT--AFT 419

QY 313 SerAspIleGluAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332
Db 420 TCCGATGTTGAAGAGATCCCGGGAACTAAGCAGAGAGATATGTGGACCTTTCCTGT 479

QY 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352
Db 480 GAAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAACAGTGGAGAAATTTGAAAAGT 539

QY 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluLysLeuAsnGluGlu 372
Db 540 CATGTAGAAAAGCTGTGATAACCAAGTTTCAAAGGTACACCTCGAAGGTTTAAATGATGAA 599

QY 373 AspValIleSerArgGlnAspHisGluGlnGlnThrGluLysLeuGluLeuGluIleGlu 392
Db 600 GATGTAATCTCAGCAGACAGCATGACAGAAACTGAAAGAACTGAGTTAGAAATTCAG 659

QY 393 ArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThr 412
Db 660 CAGTGTAAGAAATGATTAAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCA 719

QY 413 CysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeu-LeuGluGluLysGluArg 432
Db 720 TATGATGATGATCC-ACTTC-CTATTACGAGACTGTATTTGTTGGAGAAAGGACCG 777

QY 432 gleuLysGluGluTyrThrLeuPheLysGluGlnLysLys 445
Db 778 TNYTCAAGAGAAATGGGCCCTTTTAAAGAACCCNAAAAA 817

RESULT 7
AAH16996
ID AAH16996 standard; cDNA; 1503 BP.
XX
AC AAH16996;
XX
DT 26-JUN-2001 (first entry)
DE Human cDNA sequence SEQ ID NO:16285.
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
```

```
XX 28-JUL-2000; 2000EP-00116126.
PF
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 16285; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC present invention
XX
SQ Sequence 1503 BP; 490 A; 269 C; 271 G; 473 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 7,728-59 Length: 1503
Score: 856.00 Matches: 198
Percent Similarity: 44.0% Conservative: 9
Best Local Similarity: 42.1% Mismatches: 11
Query Match: 27.0% Indels: 253
Db: 4 Gaps: 1

US-10-644-084-2 (1-615) x AAH16996 (1-1503)
```

```
QY 1 MetGlyAspTrpMetThrValThrAspProValLeuLysGluAsnLysAsnLeuSer 20
Db 97 ATGGAGATTTGGATGACTGTACAGATCCAGGTCTGTCTTTCAGAAAGCAAACTACTCT 156

QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40
Db 157 CAATATACCTCAGAAACAAGATGCTCCATCAAGTTTATATCTACAGCAAGTGTATGT 216

QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60
Db 217 TCTTCAATACCTTTATCGAAAAAATGTGCACAGTTTTTTTCAGTGCCTTCTGCACAGAAAT 276

QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGln----- 71
Db 277 AATATTGAACAGAGTATCTCATATCTTGATCA-GGTAACATGTTTTTTGTAAAAAACAGTA 335
```

```
QY 71 ----- 71
Db 336 GCATAAATCATGTTATATATAGAAATGTATCAAAATTTATGAGAAAAACATTTGTCAACA 395
QY 71 ----- 71
Db 336 CAAGTTTGTGTTATGTAGTGCCCTGAGACTTAGAATTTTCAGAGAACCTTGGTGATTA 455
QY 71 ----- 71
Db 456 TCTCATCTAGAGATAGCAAGTCTAACTTTTTTTGTGTGAGCACCCTGGCCTCCTCATTTGT 515
QY 71 ----- 71
Db 516 CCCAGGGCTCTCCACTGAGAGTCCAGCCATCACTTTATGATGTCAGAAACCCAAAGTGC 575
QY 71 ----- 71
Db 576 CTTGCTCCATGTCACAGATTTATAGCAGAGACTAAAGAGCCCGTCTCTTAACCTCTTT 635
QY 71 ----- 71
Db 636 CTCCTTTGGGATTAATGACCACTTTCTTAACATATTTCAAGATAGTTTTCTATTTTCTAAG 695
QY 71 ----- 71
Db 696 GAAAGATTTCTATCTTTTGAATTTTTTTTACTAAGTTGCCAGTAACCAACCTTTTAAA 755
QY 71 ----- 71
Db 756 ATAATTCACATTTATTTTAAATTTTTCAGAAAAATTATATATATTTGTCTATGTTCAACAT 815
QY 71 ----- 71
Db 816 AATGTTTCATAATACACATTTGTGAAATGGCTCAGTGTGCATACCTCCCATCTATATCA 875
QY 71 ----- 71
Db 876 TTTTGTGGTGAGAAATTTACTTGCTCAGCAATTTTCAGGAATATAGTAGTAGTTATTAG 935
QY 71 ----- 71
Db 936 CTGAGGTCACCGTGTGTACAATAGATCTATTAAACTTATTCTTCTATCTAACTGAAAG 995
QY 71 ----- 71
Db 996 TTTGTATCCTCTGACCAGTATCTCTCAACTGTAAACCACTTTAAGAAAAACATTTTCTCC 1055
QY 72 -----GluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGlu 89
Db 1056 TACCAGGAATGCACTACTTTTGGTTTCTTCATTTATGAGATCCAAAGGTAAAGAG 1115
QY 90 AlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeu 109
Db 1116 ACNAAAGACAGAGTTAAATATAGTAGCTGTACTATAATTTGTATGAATGAGCTGCTGTGCTT 1175
QY 110 GlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysGlySer 129
Db 1176 CAGCGGAAGAACCTTCTAGCTCAGGAAAATGTGGAGACACAGAAATTTGAAGCTGGGAAGT 1235
QY 130 AspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArg 149
Db 1236 GATATGGACCATCTACAGAGCTGTCTCAAAACTTTAAGGAACAACCTGGNAACCTCCAGG 1295
QY 150 ArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeu 169
Db 1296 AGGGAATGATTTGGGCTTCAGGAAAGACAGACAGTGTACAATGTAAAGACAGGAATTTG 1355
QY 170 HisGlnLeuLeuLysAsnGluLysAspGluValGln-LysLeuGlnAsnIleIleAlaSer 189
Db 1356 CATCAGCTACTAAAGAATGAGAAAGATGAGGTGCAAAAAAATTTACAAAATATATCATTTGCAAG 1415
```

```
QY 189 rArgAlaThrGlnTyrAsnHisAspValLysGluArgGluTyrAsnLysLeuLys 209
Db 1416 TCGAGCTACTCAGTATATATCATGATGATCAAGAGAAAGACGCGTGAATATATAAAGTCAA 1475
QY 209 sGluArgLeuHisGlnLeuValMetAsn 218
Db 1476 GGAACGCTCATCACTCAACTTGTATTGAAC 1503
RESULT 8
ID ADQ78551 standard; DNA; 668 BP.
AC ADQ78551;
DT 21-OCT-2004 (first entry)
XX Novel canine microarray-related DNA sequence SeqID10110.
DE canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX
OS Canis familiaris.
SS
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
XX 05-MAY-2003; 2003WO-US013853.
XX
XX 03-MAY-2002; 2002US-0377240P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PFIZ ) PFIZER PROD INC.
XX
XX Diggins JC, Potter M, Wei T;
XX WPI; 2004-561890/54.
XX
XX New isolated nucleic acid molecule, useful for drug screening and
XX toxicity assays or for assessing the impact, including toxicity, of a
XX compound, pharmaceutical agent or environmental pollutant on a cell or
XX living organism.
XX
XX Claim 1; SEQ ID NO 10110; 41bp; English.
XX
XX This invention is related to a novel isolated canine nucleic acid
XX sequences and the construction of canine microarrays containing a
XX significant portion of the canine genome. The isolated canine nucleic
XX acid sequences of the invention may be useful for drug screening and
XX toxicity assays. The invention is therefore useful for assessing the
XX impact, including toxicity, of a compound, pharmaceutical agent or
XX environmental pollutant on a cell or living organism. The methods are
XX useful for detecting genes that are up- or down-regulated in canines in a
XX disease state. The sequences are useful as diagnostic agents or markers
XX to detect a cellular response in a sample individually or as part of a
XX gene expression profile. It is also useful as a target for agents that
XX modulate gene expression or activity. The database is useful for
XX producing electronic Northernblots that allow the user to determine the cell
XX type or tissue in which a given gene is expressed and to allow
XX determination of the abundance or expression level of a given gene in a
XX particular tissue or cell. The methods are useful for determining the
XX similarity of a toxic response to one or more individual compounds. The
XX methods are useful for predicting at least one toxic response or the
XX likelihood that a compound or test agent will induce various specific
XX pathologies such as those of the liver (liver necrosis, fatty liver
XX disease, protein adduct formation or hepatitis), those of the kidney,
XX heart, brain or testes, or other pathologies associated with at least one
XX of the toxins. The methods are also useful for predicting or elucidating
XX the potential cellular pathways influenced, induced or modulated by the
XX compound or test agent due to the similarity of the expression profile
XX compared to the profile induced by a known toxin. The present sequence is
```



CC that of a canine DNA sequence which was claimed for use during the  
 CC production of a canine microarray of the invention. Note: The sequence  
 CC data for this patent does not form part of the printed specification but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 668 BP; 205 A; 148 C; 141 G; 157 T; 0 U; 17 Other;

## Alignment Scores:

Pred. No.: 3,93e-59 Length: 668  
 Score: 854.00 Matches: 160  
 Percent Similarity: 86.3% Conservative: 17  
 Best Local Similarity: 78.0% Mismatches: 28  
 Query Match: 27.0% Indels: 0  
 DB: 13 Gaps: 0

US-10-644-084-2 (1-615) x ADQ78551 (1-668)

QY 411 ThrThrCysAspAspThrThrSerLeuLeuArgAspCysThrLeuLeuGluGluLys 430  
 DB 1 ACAGCATGCGATGATGACANACGCTCTTCTCGGAGACTGTACTTGCNNNNNAGAA 60  
 QY 431 GluArgLeuLysGluGluThrThrLeuPheLysGluLysLysAsnPheGluArgGlu 450  
 DB 61 NNGNNCTCAANNAGATGTTNCCTGTTTAAAGAGCAAAAAGAGATTTCGAGAGGA 120  
 QY 451 ArgArgSerPheThrGluAlaAlaAlaArgLeuGluLysLysAlaPheGluGlu 470  
 DB 121 AGACGAAGCTTTACAGAGAGCCATTCGCTTAGGATTGGAGAAAGCGTTTGAAGA 180  
 QY 471 GluArgAlaSerTrpValLysGlnPheLeuAsnMetThrAsnPheAspHisGlnAn 490  
 DB 181 GAAAGAGCCAGTGTGTTAAAGCAACAGTTTTTAAACATGACTTACCTTTGACCACCA 240  
 QY 491 SerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAspProAsnLeuIle 510  
 DB 241 TCAGAAAATATCAAACTTTTCAGTGCTTCTCAGGAGTTCGTGCGGACAGTCTCTCA 300  
 QY 511 ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys 530  
 DB 301 GTGCACCTCGAGGCCAGCGCAAAAGAGCCCTCAGTGTCTTAACGGGACTCCAGTTGC 360  
 QY 531 ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr 550  
 DB 361 ACCTTAACTTACTTAAGTCTTCTCTGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 420  
 QY 551 HisSerCysValSerGluHisSerSerValLeuAsnIleThrProGluGluSer 570  
 DB 421 CGGTCTCGGCATCTGAGCATAGTTCATCATGTACTGAATATTAATCTCTGAAGAACT 480  
 QY 571 LysProSerGluValAlaArgGluSerThrAspGlnLysTrpSerValGlnSerArgPro 590  
 DB 481 AAACCAATCAGTTCGCGAGAGAGAGTACAAATCAGAAATGAGCATGGCATCAAGACT 540  
 QY 591 SerSerArgGluGlyCysThrSerGlyCysSerSerAlaPheArgSerAlaHisGlyAsp 610  
 DB 541 GGATCAGAGAGAGGGTGTACAGTGGATGCTCTTTCACCTACACAAACTCCCATGTAGAA 600  
 QY 611 ArgAspAspLeuPro 615  
 DB 601 AAAGATGACTTACTT 615

## RESULT 9

AAH08911/c

ID AAH08911 standard; cDNA; 563 BP.

XX

AC AAH08911;

XX

DT 26-JUN-2001 (first entry)

DE Human cDNA clone (3'-primer) SEQ ID NO:5746.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX

OS Homo sapiens.

XX EP1074617-A2.

XX

PD 07-FEB-2001.

XX

PF 28-JUL-2000; 2000EP-00116126.

XX

PR 29-JUL-1999; 99JP-00248036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX

(HELI-) HELIX RES INST.

XX

Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

WPI; 2001-318749/34.

XX

PR Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or

PT diagnosis of the abnormality of the proteins encoded by the full-length

PT cDNAs.

XX

PS Claim 3; SEQ ID NO 5746; 2537pp + Sequence Listing; English.

XX

CC The present invention describes primer sets for synthesizing 5602 full-

CC length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dr primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

CC oligonucleotides, all of which are used in the exemplification of the

CC present invention

XX

SQ Sequence 563 BP; 144 A; 115 C; 97 G; 199 T; 0 U; 8 Other;

XX

Alignment Scores:

Pred. No.: 4.28e-48 Length: 563

Score: 715.00 Matches: 145

Percent Similarity: 94.9% Conservative: 4

Best Local Similarity: 92.4% Mismatches: 5

Query Match: 22.6% Indels: 3

DB: 4 Gaps: 1

US-10-644-084-2 (1-615) x AAH08911 (1-563)

QY 63 GluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyr 82

DB 463 GAANCAATTTCTCTCTAC-----CAGGAATGACTCTTTTGGTTTCTTCATTATAT 410

QY 83 GluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCys 102

DB 409 GAAGAATCCAAAGGTAAAGAGACAAAGAGAGATTAAATATAGTAGTGTACTTAATTGT 350

QY 103 MetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThr 122

|||||  
349 ATGATGAGCTGCTTGTCTTCAGCGGAAGAACCTTCTAGCTCAGAAAATGTGGAGACA 290  
QY 123 GlnAenLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyAlaLysLeuLys 142  
|||||  
289 CAGAAATTTGAAGCTGGGAAGTGATATGACCATCTACAGAGCTGCTACTCAAAACTTAAG 230  
QY 143 GluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeu 162  
|||||  
229 GAACAACCTGGAAACCTCCAGGAGGGAATGATTGGGCTTCAGAAAGAGACAGACAGTTA 170  
QY 163 GlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluValGln-Ly 182  
|||||  
169 CAATGTAGACAGAGAAATTTGCATCAGCTACTAAGAAATGAGAAAGATGAGGTGCAAAA 110  
QY 182 sleuGlnAenIlelleAlaSerArgAlaThrGlnTyrAenHisAspValLysArgLysGI 202  
|||||  
109 ATTCAAAAATATCATTTGCAAGTCGAGCTACTCAGTATAATCATGATATGAAGAGAAAAGA 50  
QY 202 uArgGluTyrAenLysLeuLysGluArgGluHisGlnLeuValMetAen 218  
|||||  
49 GCGTGAATATAATAAACTGAAGGAACGCTACATCAACTTGTATTGAAC 1

## RESULT 10

AAV86670

ID AAV86670 standard; cDNA; 498 BP.

AC AAV86670;

XX 27-APR-1999 (first entry)

DE EST clone BG250.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;

KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;

KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;

KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX WO9845435-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US006954.

XX 10-APR-1997; 97US-00835913.

XX (GENY ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;

XX Spaulding V, Agoetino MJ;

XX WPI; 1999-070076/06.

XX New polynucleotides encoding human secreted proteins - derived from e.g.

XX human blood, kidney, foetal lung, placenta, testes, brain, ovary,

XX pituitary, retina and colon cDNA libraries.

XX Claim 1; Page 311; 633pp; English.

XX This sequence represents an expressed sequence tag (EST), and is a

XX polynucleotide of the invention. The polynucleotides of the invention are

XX all secreted EST sequences isolated from a variety of human tissue

XX sources. The EST sequences and proteins encoded by them are predicted to

XX have useful biological activities which would make them suitable for

XX treating, preventing or ameliorating medical conditions in humans and

XX animals, although no supporting data is given. Suggested activities

XX include nutritional activity, immune stimulating or suppressing activity,

XX haematopoiesis regulating activity, tissue growth activity,

XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory

XX activity, cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The EST sequences are also stated to be useful for gene therapy  
XX Sequence 498 BP; 186 A; 71 C; 120 G; 115 T; 0 U; 6 Other;  
SQ

## Alignment Scores:

Pred. No.: 1,11e-44 Length: 498  
Score: 671.50 Matches: 136  
Percent Similarity: 90.1% Conservative: 10  
Best Local Similarity: 84.0% Mismatches: 15  
Query Match: 21.2% Indels: 2  
DB: 2 Gaps: 1

US-10-644-084-2 (1-615) x AAV86670 (1-498)

QY 224 IleAlaMetAspValLeuAenTyrValGlyArgAlaAspGlyLysArgGlySerTrpArg 243

Db 16 CTGCTATGGACATTTTGAATTATGTCGGGAGAGCTGATGGANAAGAGGCTCCNGGAGG 75

QY 244 ThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAenAspTyr 263

Db 76 ACTGGTTAACTGAAGCNGGGAATGA-GATGAATGTATAAAATTTCTTGAATGATTAT 134

QY 264 GluTyrArgGlnLysGlnIleLeuMetGluAenAlaGluLeuLysLysValLeuGlnGln 283

Db 135 GAATATCGTCAGAAACAAATCNTAATGAAAAATGCAGAACTTAAGAGGTTCTTCAACAA 194

QY 284 MetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAla 303

Db 195 ATGAAAAAGGAAATGATTTCCTTTCTCCCAAAAGAGAAACCTCGAGAAAGATGA 254

QY 304 GluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSer 323

Db 255 GATGATAGTACAGGAACTGTT--ATTTCGATGTTGAAGAAGATGCCNGGGAACCTAAGC 311

QY 324 ArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIle 343

Db 312 AGAGAGAGTATGTGGACCTTTCTCTGTGAAACTGTGAGAGAGCAGCTTACAAACAGCATC 371

QY 344 ArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAenGlnAlaSerLys 363

Db 372 AGAAAAACAGTGGAGAAATTTTGAAGAAGTCATGTAGAAAAGCTTTGATAACCAAGTTTCAAG 431

QY 364 ValHisSerGluGlyLeuAenGluGluAspValIleSerArgGlnAspHisGluGlnGlu 383

Db 432 GTACACCTGGAGAGGTTTATATGATGAAGATGTATCTCACCAGCAAGACCATGAACAGAA 491

QY 384 ThrGlu 385

Db 492 ACTGAA 497

## RESULT 11

AAF66702

ID AAF66702 standard; cDNA; 403 BP.

XX AAF66702;

AC AAF66702;

XX 09-APR-2001 (first entry)

XX Novel human polynucleotide, SEQ ID NO: 2458.

XX Human; cytostatic; gene therapy; colon cancer; prostate cancer;

XX breast cancer; lung cancer; cancer detection; ss.

XX Homo sapiens.

XX WO200102568-A2.

XX 11-JAN-2001.

XX 30-JUN-2000; 2000WO-US018374.

XX 02-JUL-1999; 99US-0142310P.

XX 02-JUL-1999; 99US-0142311P.

```
XX (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
PI Kassam A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
PI Dmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
DR WPI; 2001-091805/10.
XX
XX Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences.
XX
XX Claim 9; Page 906; 1046pp; English.
XX
XX The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia
XX
SQ Sequence 403 BP; 162 A; 57 C; 96 G; 88 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 4,46e-43 Length: 403
Score: 650.00 Matches: 124
Percent Similarity: 97.7% Conservative: 4
Best Local Similarity: 94.7% Mismatches: 3
Query Match: 20.5% Indels: 0
DB: 5 Gaps: 0
US-10-644-084-2 (1-615) x AAF66702 (1-403)
QY 132 AspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgGlu 151
Db 10 GACCATCTACAGAGCTGCTACTCAAAACTTATGGACAACTGGAACTCTCAGGAGGAA 69
QY 152 MetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGln 171
Db 70 ATGATTGGGCTTCAGGAAAGAGACAGACAGTACATGTAAGAACAGGAATTTGCATCAG 129
QY 172 LeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAla 191
Db 130 CTACTAAAGAAATGAGAAAGATGAGGTGGCAAAATACAAAATATCATTCAGTCGAGCT 189
QY 192 ThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArg 211
Db 190 ACTCAGTATAATCATGATATGAGAGAAAGAGCGGTGAATATAATAACTGAAGNACGT 249
QY 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyr 231
Db 250 CTACATCAACTGTTTATGAACAAGAAAGATAAGAAAATAGCTATGACATTTTGAATTAT 309
QY 232 ValGlyArgAlaAspGlyLysArgGlySerTTrpArgThrAspLysThrGluAlaArgAsn 251
Db 310 GTCGGAGAGCTGATGGAAAAAGAGGCTCTCGAGGACTGGTAAAACTGAAGCCAGGAAT 369
QY 252 GluAspGluMetTyrLysIleLeuLeuAsnAsp 262
Db 370 GAAGATGAAATGTATAAAATTTCTTTGATGAT 402
```

```
RESULT 12
AAC03006
ID AAC03006 standard; cDNA; 428 BP.
XX
XX AAC03006;
AC AAC03006;
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3004.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-00200610.
XX
PR 26-FEB-1999; 99US-0122487P.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
P-PSDB; AAG03000.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 1; SEQ ID NO 3004; 71pp + Sequence Listing; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX
SQ Sequence 428 BP; 170 A; 69 C; 94 G; 95 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 6,99e-42 Length: 428
Score: 635.50 Matches: 125
Percent Similarity: 94.4% Conservative: 10
Best Local Similarity: 87.4% Mismatches: 7
Query Match: 20.1% Indels: 1
DB: 3 Gaps: 1
US-10-644-084-2 (1-615) x AAC03006 (1-428)
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLys 286
Db 1 CAGAAACAAATCTTAATGGAANAATGCAGACTTAAGAAGGTTCTTCAACAATGAAAAG 60
QY 287 GluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGly 306
Db 61 GAAATGATTTCTCTTTCTTCTCCCAAAAGAAACCTAGAGAAAGAGTAGATGATAGT 120
QY 307 ThrGlyThrValAlaIleSerAspIleGluAspSerGlyGluLeuSerArgAspSer 326
|||||
|||||
|||||
```

Db 121 ACAGGAAGTGT--ATTCCGATGTTGAAGAAGATGCCGGGAACTAAGCAGAGAGT 177  
 Qy 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346  
 Db 178 ATGTGGGACCTTCTCTGTGAAACTGTGAGAGAGCAGCTTACAAACAGATCAGAAACAG 237  
 Qy 347 TTPArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366  
 Db 238 TGGAGAATTTTCAAAAGTCATGTAGAAAGCTTGATAACCAAGTTTCAAGGTACACCTG 297  
 Qy 367 GluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLys 386  
 Db 298 GAAGGTTTAAATGAAGATGAATCTCACGACAGACCATGAACAAGAACTGAAAAA 357  
 Qy 387 LeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuGln 406  
 Db 358 CTCGAGTTAGAAATTCAGCAGGTGAAGAAGATGATTAAGAACTCAGCAACAGCTTTTACAG 417  
 Qy 407 GlnGlnLeu 409  
 Db 418 CAGCAGCTC 426  
 RESULT 13  
 ID ADX36698  
 AC ADX36698 standard; cDNA; 1691 BP.  
 AC ADX36698;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Plant full length insert polynucleotide seqid 19518.  
 XX  
 KW plant protectant; plant growth regulant; gene therapy; plant;  
 KW recombinant DNA construct; physical array; plant breeding marker;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
 KW growth rate; cell cycle pathway; disease resistance;  
 KW galactomannan production; lignin production; plant growth regulator;  
 KW yield; plant growth; plant development; seed oil; protein yield;  
 KW protein content; gene; ss.  
 XX  
 OS Unidentified.  
 XX  
 PN US2004034888-A1.  
 XX  
 PD 19-FEB-2004.  
 XX  
 XX 28-APR-2003; 2003US-00425114.  
 PF  
 PR 06-MAY-1999; 99US-00304517.  
 PR 05-NOV-2001; 2001US-00985678.  
 XX  
 XX (LIU/) LIU J.  
 PA (ZHOU/) ZHOU Y.  
 PA (KOVA/) KOVALIC D K.  
 PA (SCRE/) SCREEN S E.  
 PA (TAB/) TABASKA J E.  
 PA (CAO/) CAO Y.  
 XX  
 PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
 XX  
 DR WPI; 2004-180133/17.  
 XX  
 PT New recombinant DNA construct, useful for improving plant tolerance to  
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
 PT pests, for conferring increased resistance to plant disease, or for  
 PT improving yield.  
 XX  
 PS Claim 1; SEQ ID NO 19518; 15pp; English.  
 XX  
 XX The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at

CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This sequence represents a plant full length insert  
 CC polynucleotide that can be used in the recombinant DNA construct of the  
 CC invention.

SQ Sequence 1691 BP; 474 A; 409 C; 411 G; 397 T; 0 U; 0 Other;

Alignment Scores:	1.08e-20	Length:	1691
Pred. No.:	380.50	Matches:	110
Score:	48.4%	Conservative:	106
Percent Similarity:	24.7%	Mismatches:	155
Best Local Similarity:	12.0%	Indels:	75
Query Match:	13	Gaps:	14

US-10-644-084-2 (1-615) x ADX36698 (1-1691)

Qy 27 LysMetSerProSerSerLeuTyr-----SerGlnGlnValLeuCysSer 41  
 Db 160 CGGATGTCCTCGTGGCGCGTTCGATCTCGGGCCTCGTCGACAGCTTG-----CAG 213  
 Qy 42 SerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsn 61  
 Db 214 CGGCACCTCGCAGCGCAGCATGACGACGACGGGGCGCTTCGCAACCGCGAGAAC 273  
 Qy 62 IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSerLeu 81  
 Db 274 CTGGAGCACTCGCGCGGTACTGTAACACAGCGCTCGTCACTTCGGATTCCCGGCA--- 330  
 Qy 82 TyrGluGlnSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsn 101  
 Db 331 -----TCTCGACCTCTTCGCCACCGATCGGTCCTCGATTGCAAGAACTTGTAAAC 381  
 Qy 102 CysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlnUserValGlu 121  
 Db 382 TGCATCTATGCTGCTGCTGACGACGACCAAGGACATTGAGTTCAGGGAGTCTACAAAT 441  
 Qy 122 ThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeu 141  
 Db 442 GACCAGCGCAGCGGATGCAATCTGATATCTCACGGCTGGAAGCTAAGATTGAGAGNAATG 501  
 Qy 142 LysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGlnUArgAspArgGln 161  
 Db 502 GATGCTCAATTAGCAGCGAAAGATCGAGAGCTAGCCACTCTTACTCGAACGAGGCCAAA 561  
 Qy 162 LeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGlnLysAspGluValGln 181  
 Db 562 AACACTGCTGCTCAAGTCTCAGATTGATTAAGCTGCAACAAAGACGTGATGAATTCAG 621  
 Qy 182 LysLeuGlnAsnIleIleAlaSerArg-----AlaThrGlnTyrAsnHisAspValLys 199  
 Db 622 AAAATG-----GTTATAGGAATTCAGCAAGTACGCCACCCAGCAAAATTCATGAATGAAG 675  
 Qy 200 ArgLysGluArgGluTyrAsnLysLysGluArgLeuLeuGlnLeuValMetAsnLys 219  
 Db 676 AAGAAAGAAAGAAATACATAAGTTGAGAAAGCTAAATCAGCTATTGATGAGAGAG 735  
 Qy 220 LysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGly 237  
 Db 736 AAAAAGGAATCATCCCGTTCTGGAATGGAGATAATGAACCTGTTACAGAAAGAACGACGA 795  
 Qy 238 LysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLys 257

[illegible]

Qy	139	aLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAs	159	XX	06-SEP-2000.	
Db	410	AAAGATGGATGCTCAATAGCTGCAAAAGATCGCAGCTGCCACATTTAGTGAACGGA	469	XX	25-FEB-2000; 2000EP-00301439.	
Qy	159	pArgGlnLeuGlnCysLeuAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspG	179	XX	25-FEB-1999; 99US-0121825P.	
Db	470	GCCCAAAACACTGCAACTTTCAAGGCTCAGATTCAACAGCTGCAACAGGAACGCGATGA	529	PR	05-MAR-1999; 99US-0123180P.	
Qy	179	uValGlnLeuLeuGlnAsnIleAlaSerArg-----AlaThrGlnTyrAsnHisAs	197	PR	09-MAR-1999; 99US-0125748P.	
Db	530	GTTCAAAAAATG-----GTTATTGGCAATCAGCAAGTACGTACCCAGCAAAATTCATGA	583	PR	23-MAR-1999; 99US-0126264P.	
Qy	197	pValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValme	217	PR	29-MAR-1999; 99US-0126785P.	
Db	584	AATGAAGAAAAAGAGAAAGAAATACATCAATTCGAGGAGAAGTTAAACCCAGGTATAAT	643	PR	01-APR-1999; 99US-0127462P.	
Qy	217	tAsnLysLysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAl	235	PR	08-APR-1999; 99US-0128234P.	
Db	644	CGAGAAAAAAGAGGAATCATACGTTTCAGGAATGGAATAATGAACCTTTGTCAGAAAGA	703	PR	16-APR-1999; 99US-0129845P.	
Qy	235	aAspGlyLysArgGlySerTyrArgThrAspLysThrGluAlaArgAsnGluAspGluMe	255	PR	19-APR-1999; 99US-0130077P.	
Db	704	AGCAGCGCAGCTGGNACTTGAGTGGAAAAAG-----AATGACAAATGATTA	751	PR	21-APR-1999; 99US-0130449P.	
Qy	255	tTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAl	275	PR	23-APR-1999; 99US-0130510P.	
Db	752	TTACAAAATGATTTGATGCTATGAGGTAAAGAACGAGAGTGTGATGCAAGAAATGC	811	PR	28-APR-1999; 99US-0130891P.	
Qy	275	aGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGl	295	PR	30-APR-1999; 99US-0132048P.	
Db	812	TGATTTGCGAGCACTATTGCGTTCGATGCAGATGATATGCGCGAGTTCCTCAATGCTCC	871	PR	04-MAY-1999; 99US-0132484P.	
Qy	295	nLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIl	315	PR	05-MAY-1999; 99US-0132485P.	
Db	872	AAATGGAGTTCCA---CACCGAGCTGTGTGGCAATGACGACAGGAGCGAGGATCTCC	928	PR	06-MAY-1999; 99US-0132486P.	
Qy	315	eGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrVa	335	PR	07-MAY-1999; 99US-0132863P.	
Db	929	TCAGTCTCCACTTGGT-----GGCAAGACGGATGCTTTGATTTGCCCTTCACATGGC	982	PR	11-MAY-1999; 99US-0134256P.	
Qy	335	lArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGl	355	PR	14-MAY-1999; 99US-0134218P.	
Db	983	TAGACACCAAGATTGAAGAGAGTTTCGCACATAAATGGCTTCATAAAGGCCGGAATGAC	1042	PR	14-MAY-1999; 99US-0134219P.	
Qy	355	uLysLeuAspAsn-----GlnAlaSerLysValHisSerGlyLeuAsnGluLysAs	373	PR	14-MAY-1999; 99US-0134221P.	
Db	1043	ACAACCTTCAAGATGCCCAAAAGGGTGCAGAAATGACATCAGAAAGCTACC-----	1091	PR	18-MAY-1999; 99US-0134941P.	
Qy	373	pValIleSerArgGlnAspHisGlnGlnGluThrGluLysLeuGluLeuGluLeGluAr	393	PR	20-MAY-1999; 99US-0135124P.	
Db	1092	-----GAGCGGGAACCTTCAA---CTTGAAGCTCAACTAGTTGA	1126	PR	21-MAY-1999; 99US-0135333P.	
Qy	393	gCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGln	407	PR	24-MAY-1999; 99US-0135629P.	
Db	1127	AGCTAGAAGCATTTATCCAGGACGAGGCATCCCTCATGTGTCGAAG	1169	PR	25-MAY-1999; 99US-0136021P.	
RESULT 15						
AAC43689						
ID AAC43689		standard; DNA; 1544 BP.				
XX						
AC						
XX						
DT						
DE						
XX						
XX						
KW						
KW						
XX						
OS						
XX						
PN						

Hybridisation assay; genetic mapping; gene expression control;  
protein identification; signal transduction pathway; metabolic; pathway;  
promoter; termination sequence; corn; ss.  
Zea mays subsp. mays.  
EP1033405-A2.

```
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142777P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 20-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.

PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157513P.
PR 06-OCT-1999; 99US-0157863P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 1,548-18 Length: 1544
Score: 353.00 Matches: 103
Percent Similarity: 48.6% Conservative: 98
Best Local Similarity: 24.9% Mismatches: 145
Query Match: 11.2% Indels: 69
DB: 3 Gaps: 12

US-10-644-084-2 (1-615) x AAC43689 (1-1544)

Qy 54 GlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeu 73
Db 185 GCGCGTTTCGCAACCGCGAGAACCTGGAGCACTCGCCCGGTACTGTAACACGAGCGTC 244
Qy 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93
Db 245 GTCACCTTCGGATTCCCGGCA-----TCTCTCGACCTCTTCGCCACCGATCCG 292
Qy 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsn 113
Db 293 GTCTCGATTGCAGNACTTGTAACATGTCATCTATGCGCTGTCGACGACGAGGAGGAC 352
Qy 114 LeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHis 133
Db 353 ATTGAGTTTCAGGAGTCTACAAATGACCAGCGGCA-SGNATGCAATCTGATATCTCACGG 411
Qy 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIle 153
Db 412 CTGGAAGCTAAGATAGAGAGATGGATGCTCAATTAGCAGCGAAAGATCGAGAGTAGCC 471
Qy 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173
Db 472 ACTCTTACTGAACSGAGGCCAAAACACTGCTGCTCTGTAAGTCTTCAGATTGATAAGCTG 531
```





GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 13, 2006, 23:33:20 ; Search time 1957 Seconds  
(without alignments)  
5792.204 Million cell updates/sec

Title: US-10-644-084-2  
Perfect score: 3165  
Sequence: 1 MGDWMTVPVLCNTENKLS.....CYSGSSAFPSAHGRDDLP 615

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 18992170 seqs, 6143817638 residues

Total number of hits satisfying chosen parameters: 37784340

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DRV=xlh  
-Q=/abes/ABSSWEB spool/US10644084/runat\_12062006\_150429\_1550/app\_query.fasta\_1  
-DB=Published Applications NA Main -OFMT=fastap -SUFFIX=p2n.rnpbm  
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext  
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs808  
-USER=US10644084 @CEN 1.1 1675 @runat\_12062006\_150429\_1550 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:  
1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
7: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
8: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
9: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
10: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
11: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
12: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US10G\_PUBCOMB.seq.\*  
13: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11A\_PUBCOMB.seq.\*  
14: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11B\_PUBCOMB.seq.\*  
15: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11C\_PUBCOMB.seq.\*  
16: /EMC\_Celerra\_SIDS3/ptodata/2/pubpna/US11D\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2720	85.9	2272	7	US-10-104-047-1497 Sequence 1497, Ap

2	2720	85.9	2272	16	US-11-072-512-1497 Sequence 1497, Ap
3	650	20.5	403	15	US-10-779-543-12377 Sequence 12377, A
4	438	13.8	161874	15	US-11-121-086-75 Sequence 75, Appl
5	380.5	12.0	1691	8	US-10-425-114-19518 Sequence 19518, A
6	379	12.0	1567	8	US-10-424-599-27277 Sequence 27277, A
7	378	11.9	1805	9	US-10-425-115-157154 Sequence 157154, A
8	361.5	11.4	3750	8	US-10-437-963-34854 Sequence 34854, A
9	334.5	10.6	1062	9	US-10-767-795-5235 Sequence 5235, Ap
10	328	10.4	513	7	US-10-029-386-5594 Sequence 5594, Ap
11	290	9.2	1199	8	US-10-425-114-10108 Sequence 10108, A
12	285	9.0	178	7	US-10-029-386-19350 Sequence 19350, A
13	234.5	7.4	5040	7	US-10-369-493-45767 Sequence 45767, A
14	233	7.4	730	8	US-10-767-701-7320 Sequence 7687, A
15	233	7.4	3444	8	US-10-437-963-77687 Sequence 77687, A
16	231.5	7.3	5373	11	US-10-932-182A-5516 Sequence 5516, Ap
17	230.5	7.1	5040	9	US-10-828-985A-10 Sequence 70555, A
18	225.5	7.1	7509	9	US-10-828-985A-10 Sequence 145, App
19	223.5	7.1	10300	3	US-09-960-253-145 Sequence 49, Appl
20	223.5	7.1	10300	9	US-10-422-524-49 Sequence 122, App
21	223.5	7.1	10300	10	US-10-947-249-122 Sequence 45972, A
22	220.5	7.0	5628	7	US-10-369-493-45972 Sequence 40, Appl
23	220.5	7.0	6386	6	US-10-098-841-40 Sequence 40754, A
24	219.5	6.9	2602	13	US-11-097-143-40754 Sequence 134, App
25	218	6.9	11677	6	US-10-082-830-134 Sequence 6, Appl
26	217.5	6.9	4470	7	US-10-205-219-6 Sequence 33485, A
27	217	6.9	3614	13	US-11-097-143-33485 Sequence 117, App
28	217	6.9	6921	3	US-09-735-705-117 Sequence 117, App
29	217	6.9	6921	3	US-09-850-716A-117 Sequence 117, App
30	217	6.9	6921	3	US-09-897-778-117 Sequence 117, App
31	217	6.9	6921	3	US-09-466-398A-117 Sequence 117, App
32	217	6.9	6921	6	US-10-007-700-117 Sequence 117, App
33	217	6.9	6921	7	US-10-117-982-117 Sequence 117, App
34	217	6.9	6921	7	US-10-313-986-117 Sequence 117, App
35	217	6.9	6921	9	US-10-775-972-117 Sequence 117, App
36	217	6.9	6921	10	US-10-922-124-117 Sequence 117, App
37	217	6.9	6921	10	US-10-623-155-117 Sequence 117, App
38	217	6.9	7045	3	US-09-919-172-28 Sequence 28, Appl
39	217	6.9	7045	9	US-10-752-986-28 Sequence 170, Appl
40	217	6.9	8684	7	US-10-205-219-170 Sequence 223, App
41	217	6.9	8684	10	US-10-756-149-223 Sequence 649, App
42	217	6.9	8684	10	US-10-287-436A-649 Sequence 219, App
43	217	6.9	8974	7	US-10-341-434-219 Sequence 229, App
44	217	6.9	8974	7	US-10-341-434-229 Sequence 17, Appl
45	217	6.9	8974	10	US-10-287-436A-17

ALIGNMENTS

RESULT 1  
US-10-104-047-1497  
; Sequence 1497, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1el full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; CURRENT FILING DATE: 2002-03-25  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1497  
; LENGTH: 2272  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-1497

Alignment Scores:				
Pred. No.:	5.3e-224	Length:	2272	
Score:	2720.00	Matches:	536	
Percent Similarity:	86.8%	Conservative:	34	
Best Local Similarity:	81.6%	Mismatches:	40	

Query Match: 85.9% Indels: 47  
DB: 7 Gaps: 3  
US-10-644-084-2 (1-615) x US-10-104-047-1497 (1-2272)

QY 1 MetGlyAspTrpMetThrValThrAspProValLeuCysThrGluAsnLysAsnLeuSer 20  
DB 224 ATGGGAGATTGGATGACTGTTACAGATCCA-----GAAAGCAAAACTATCTCT 271

QY 21 GlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCys 40  
DB 272 CAATATACCTCGAAACAAAGATGCTCCATCAAGTTTATATCTCACAGCAAGTGTATGT 331

QY 41 SerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu 60  
DB 332 TCTTCATACCTTATCGAAAATGTGCACAGTCTTTTCAGTGCCTCTGTCACAGAGAT 391

QY 61 AsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPheProSer 80  
DB 392 AATATTGAACACAGATATCTCATATCTTGATCAGGAATTGACTACTTTTGGTTTCTCTCA 451

QY 81 LeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeu 100  
DB 452 TTATATTGAAGAAATCCAAAGGTAAAGACAAAGAGAGAGTTAAATATATAGTAGCTACTA 511

QY 101 AsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerVal 120  
DB 512 AATTGTATGAATGAGCTGCTGTGTGCTTCAGCGGAAGAACCTCTAGCTCAGGAAATGTG 571

QY 121 GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLys 140  
DB 572 GAGACACAGAAATTGAAGCCGGAGTAGATATGGACCCTCTACAGAGCTGCTACTCAAAA 631

QY 141 LeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArg 160  
DB 632 CTTAAGGAACAACCTGGAACCTCCAGAGGGGAAATGATTGGGCTTCAGGAAAGAGACAGA 691

QY 161 GlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGluLysAspGluVal 180  
DB 692 CAGTTACATGTATAGACAGGAATTGTCATCAGCTACTTAAAGATGAGNAAGATGAGGTG 751

QY 181 GlnLysLeuGlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArg 200  
DB 752 CAAMAATTACAAATATCATTCGACGTCGAGCTACTCAGTAAATCATGATATGAAGAGA 811

QY 201 LysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLys 220  
DB 812 AAAGAGCGTGAATATATAAATGAAGAACGCTTACATCAACTTGTATGAACAGAAA 871

QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
DB 872 GATAAGAAATAGCTATGACATTTTGAATTATGTCGGAGAGCTGATGAGNAAGAGGC 931

QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeu 260  
DB 932 TCTCGAGGACTGGTAAAACTGAAGCAGGAATGAAGATGAATGATAAAATTCCTTTG 991

QY 261 AsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysVal 280  
DB 992 AATGATTATGAATATCGTCAGAAAACAATCCCTAATGGAAAAATGCAGAACTTAAGAGGTT 1051

QY 281 LeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArg 300  
DB 1052 CTTCAACAATGAAAAGGAATGATTCTCTCTCTCTCCCAAAAAGAGAAACCTAGA 1111

QY 301 GluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGly 320  
DB 1112 GAAAGAGTAGATAGTACAGGAACCTGTT--ATTTCGATGTGAAGAGATGCCGG 1168

QY 321 GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThr 340  
DB 1169 GAACTAAGCAGAGAGATGTGTGGGACCTTTCTCTGTGAACCTGTGAGAGAGCAGCTTACA 1228

QY 341 AsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGln 360  
DB 1229 AACAGCATCAGAAAACAGTGGAGAAATTTTGAAGTGCATGTAGAAAAGCTTGTATACCAA 1288

QY 361 AlaSerLysValHisSerGluGlyLeuAsnGluAspValIleSerArgGlnAspHis 380  
DB 1289 GTTTCAAAGGTACACCTCGAAGGTTTAAATGATGAAGATGTAATCTCAGCAAGACCAT 1348

QY 381 GluGlnGluThrGluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAla 400  
DB 1349 GAACAAGAAACTGAAAACCTCGAGTTAGAAATTCAGCAGTGTAAAGAAATGATTAACACT 1408

QY 401 GlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeu 420  
DB 1409 CAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGATGATGATACACACTTCACTA 1468

QY 421 LeuArgAspCysTyrLeuLeuGluGluLysGluArgLeuLysGluGluTrpThrLeuPhe 440  
DB 1469 TTACGAGACTGTATTTTGTGGAAAGAAAGAACGCTCTCAAAGAAAGAAATGGTCCCTTTT 1528

QY 441 LysGluGlnLysLysAsnPheGluArgGluArgArgSerPheThrGluAlaIleArg 460  
DB 1529 AAAGACGAGAAAAGAAATTTTGAAGGAGAGACGAAGCTTTACAGAAGCCGCTATTGCG 1588

QY 461 LeuGlyLeuGlu----- 464  
DB 1589 CTGGGATTGGAGATTGGCTTTCTCTGCCAAGTGTTTTACACAGAAATCTCTTTTGGGT 1648

QY 464 ----- 464

DB 1649 ATCCATCATTCATCCACGGTCAGTAGTGTGGTGTGTCAGAAGACAGTGCACCAAGACTAT 1708

QY 465 -----ArgLysAlaPheGluGluGluArgAlaSerTrpValLysGlnG 479

DB 1709 ATTTGCTGTTTAAAAAGAAAGGCATTTGAAGAGAAAGAGCAGCTTGGTTAAAGCAGCA 1768

QY 479 nPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAl 499  
DB 1769 GTTTCTAAATATGACTACTTTTGACCACCAAGAACTCAGAAAAATGTGAAACTTTTCAGTGC 1828

QY 499 aPheSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysL 519  
DB 1829 CTTCTCAGGAAGTTCTGATTGGGCAATCTTATAGTCACCTCGAGGCGAGCCGCAAAAGAA 1888

QY 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539  
DB 1889 GCCTCAGCTGTGTCTAATGGGTCTCCAGTTTGCATGTCTAACTTACTAAATCTCTTCC 1948

QY 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559  
DB 1949 TGCTTCACTTCCACTTCAGACTTTTGCCACACAGCTTCTCTGCATATCTGAACATAGTTC 2008

QY 559 rLleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579  
DB 2009 AATCAATGTACTGAATATAACTGCTGAAGAAATTAACCAATCAGGTGTGGAGGAACG 2068

QY 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGl 599  
DB 2069 TACAATCAANAATGGAGTGTGGCGTCNAGACCTGGATCAGCAGGAGGTTGCTATAGTGG 2128

QY 599 yCysSerSerAlaPheArgSerAlaHisGlyAspAspLeuPro 615  
DB 2129 ATGCTCTTGAGCTACACAAATTCATGTAGAAAAAGATGACTTACCT 2177

## RESULT 2

US-11-072-512-1497  
; Sequence 1497, Application US/11072512  
; Publication No. US20060029945A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI



Db 1769 GTTCTTAATATGACTACCTTTGACCACCAGAACTCAGAAATGTGAAACTTTTCAGTGC 1828  
Qy 499 aPeSerGlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnIleVal 519  
Db 1829 CTTCTCAGGAAGTTCGTATGGGACAACTTTATAGTGCACTCGAGCAGCCGCAAAAGAA 1888  
Qy 519 sLeuHisSerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuPr 539  
Db 1889 GCCTCACAGTGTCTAATGGGTCTCCAGTTTGCATGCTCTAACTTACTTAATCTCTTCC 1948  
Qy 539 oAlaSerProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSe 559  
Db 1949 TGCTTCACCTTCACCTTCAGACTTTTGCAGACACAGTTCCTGCATATCTGAACATAGTTC 2008  
Qy 559 rIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSe 579  
Db 2009 AATCAATGACTGAATATACTGCTGAAGAAATTAACCAATCAGTTGGAGGAGACG 2068  
Qy 579 rThrAspGlnLysTrpSerValGlnSerArgProSerSerArgGluGlyCysTrpSerGl 599  
Db 2069 TACAATCAAAATGGAGTGTGGCGTCAAGACCTGGATCACAGGAAGGTTGCTATAGTGG 2128  
Qy 599 yCysSerSerAlaPheArgSerAlaHisGlyAspArgAspLeuPro 615  
Db 2129 ATGCTCTTTCAGCTACACAAATTTCTCATGTAGAAAAAGATGACTTACCT 2177

RESULT 3  
US-10-779-543-12377  
; Sequence 12377, Application US/10779543  
; Publication No. US20050227917A1  
; GENERAL INFORMATION:  
; APPLICANT: Williams et al  
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED  
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II  
; FILE REFERENCE: 2300-21302  
; CURRENT APPLICATION NUMBER: US/10/779,543  
; CURRENT FILING DATE: 2004-02-12  
; PRIOR APPLICATION NUMBER: 10/076,555  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 09/217,471  
; PRIOR FILING DATE: 1998-12-21  
; PRIOR APPLICATION NUMBER: 60/068,755  
; PRIOR FILING DATE: 1997-12-23  
; PRIOR APPLICATION NUMBER: 60/080,664  
; PRIOR FILING DATE: 1998-04-03  
; PRIOR APPLICATION NUMBER: 60/105,234  
; PRIOR FILING DATE: 1998-10-21  
; PRIOR APPLICATION NUMBER: 09/297,648  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: PCT/US99/01619  
; PRIOR FILING DATE: 1999-01-28  
; PRIOR APPLICATION NUMBER: 60/072,910  
; PRIOR FILING DATE: 1998-01-28  
; PRIOR APPLICATION NUMBER: 60/075,954  
; PRIOR FILING DATE: 1998-02-24  
; PRIOR APPLICATION NUMBER: 60/080,114  
; PRIOR FILING DATE: 1998-03-31  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 23767  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12377  
; LENGTH: 403  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-779-543-12377  
Alignment Scores:  
Pred. No.: 2,02e-46 Length: 403  
Score: 650.00 Matches: 124  
Percent Similarity: 97.7% Conservative: 4  
Best Local Similarity: 94.7% Mismatches: 3  
Query Match: 20.5% Indels: 0  
DB: 10 Gaps: 0

US-10-644-084-2 (1-615) x US-10-779-543-12377 (1-403)  
Qy 132 AspHisLeuGlnSerCysTrpAlaLysLeuLysGluGlnLeuGluThrSerArgArgGlu 151  
Db 10 GACCATCTACAGAGCTGCTACTCAAACTTATGGAAACAACCTCGAAACCTCCAGAGGGAA 69  
Qy 152 MetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGln 171  
Db 70 ATGATTGGGCTTCAGGAAAGAGACAGACAGATTACAATGTAAAGAACAGGAATTTGCATCAG 129  
Qy 172 LeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArgAla 191  
Db 130 CTACTAAAGAATAGAAAAGATGAGGTGCAAAATTTACAAAATATCATTTGCAAGTCGAGCT 189  
Qy 192 ThrGlnTrpAsnHisAspValLysArgLysGluArgGluGluTrpAsnLysLeuLysGluArg 211  
Db 190 ACTCAGTATAATCATGATATGAAGAGAAAGAGCGTGAAATATAATAAATCTGAAGAACGT 249  
Qy 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTrp 231  
Db 250 CTACATCAACTTGTATTGACACAGAAAGATAAGAAAATAGCTATCGACATTTTGAATTAT 309  
Qy 232 ValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsn 251  
Db 310 GTCGGGAGAGCTGATGGAAAAAGAGGCTCTCGGAGGACTGGTTAAAACTGAAGCCAGGAAT 369  
Qy 252 GluAspGluMetTrpLysIleLeuLeuAsnAsp 262  
Db 370 GAAGATGAATGTATATAAAATTTCTTTGAATGAT 402  
RESULT 4  
US-11-121-086-75/c  
; Sequence 75, Application US/111121086  
; Publication No. US20050266459A1  
; GENERAL INFORMATION:  
; APPLICANT: POULSEN, TIM S.  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES  
; FILE REFERENCE: 09138.6000-00000  
; CURRENT APPLICATION NUMBER: US/11/121,086  
; CURRENT FILING DATE: 2005-05-04  
; PRIOR APPLICATION NUMBER: 60/567,570  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 75  
; LENGTH: 161874  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-121-086-75  
Alignment Scores:  
Pred. No.: 1.27e-24 Length: 161874  
Score: 438.00 Matches: 114  
Percent Similarity: 32.0% Conservative: 10  
Best Local Similarity: 29.5% Mismatches: 12  
Query Match: 13.8% Indels: 252  
DB: 15 Gaps: 1  
US-10-644-084-2 (1-615) x US-11-121-086-75 (1-161874)  
Qy 10 ProValLeuCysThrGluAsnLysAsnLeuSerGlnTrpSerGluThrLysMetSer 29  
Db 57511 CCTTTTATCTTTCAGAAAGCAAACTATCTCTCAATATACCTCAGAAACAAGATGTCT 57452  
Qy 30 ProSerSerLeuTrpSerGlnValLeuCysSerSerValProLeuSerLysAsnVal 49  
Db 57451 CCATCAAGTTTATCTACTCAGCAGAGTGTCTGTCTTCAATACCTTTTATCGAAAAATGTG 57392  
Qy 50 HisGlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerLysTrpLeu 69  
Db 57391 CACAGTTTTTTCAGTGCCTTCTGCACAGAAGATAATATTTGAACAGAGTATCTCATATCTT 57332

```
QY 70 AspGln----- 71
|||
Db 57331 GATCA-GGTAACATGTTTTTGTAAACACAGTAGCATAAATCATGTATATATAAGAATG 57273
QY 71 ----- 71
Db 57272 TATCAAAATTATGAGAAACAAATTGTGCAACACAAAGTTTGTGTTATGATAGTCCCT 57213
QY 71 ----- 71
Db 57212 GAGACTTAGATTTTCAGAGAACCTTGGTGATTATCTCATCTAGAGATAGCAAGTCTAACT 57153
QY 71 ----- 71
Db 57152 TTTTGTGTGGAGCACCTTGGGCTCTCATTTGTCAGGGGCTCTCCACTGAGAGTCCA 57093
QY 71 ----- 71
Db 57092 GCCATCACTTATGGATGCAGAAACCAAGTGCCTTGTCCATGTCACAGATTATAGCA 57033
QY 71 ----- 71
Db 57032 GAGACTAAAGAGCCCGCTCTTAACCTTTGTCTTTGGGATATTGACCAATTTCTT 56973
QY 71 ----- 71
Db 56972 AACTATTCAAGATAGTTTTTCTATTCTTAAGAAAGATTCTCTATCTTTTGATTTTTT 56913
QY 71 ----- 71
Db 56912 TTTACTAAGTTGCCAGTAACCAACCTTTTAAATAATTCACATTTATTTTAAATTTTG 56853
QY 71 ----- 71
Db 56852 ACAAAATTTATATATTGTCATGTTCAACATGATGTTTTCATAATACATTTGTGAAT 56793
QY 71 ----- 71
Db 56792 GGCTCAGTGTGCATACCTCCCATACTTATATCATTTTTTTGTGTGAGAAATTTACTTGC 56733
QY 71 ----- 71
Db 56732 AGCAATTTTCAAGNATATAGTACGTAGTTATTAGTCAAGGTACCGTGTGTACATAGAT 56673
QY 71 ----- 71
Db 56672 CTATTAAACTTATTTCTCTATCTAACTGAAAGTTTGTATCTCTGACCAGTATCTCTC 56613
QY 72 ----- 72
Db 56612 AACTGTAAACAACTTTAAGAAACAAATTTTCTCTACAGAAATTCGACTACTTTTGTGTTT 56553
QY 79 ProSerLeuTyr:GluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla 98
|||
Db 56552 CTTTCATTATATGAAGATCCAAAGTAAAGAGACAAAGAGAGAGTTAAATATAGTAGCT 56493
QY 99 ValLeuAsnCyMetAsnGluLeuValLeuValLeuValLeuValLeuValLeuValLeu 118
|||
Db 56492 GTACTAAATTGTATGAATCAGCTGCTTGTGCTTTCAGCGAAGAACCTTCTAGCTCAGAA 56433
QY 119 SerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyr 138
|||
Db 56432 AATGTGGAGACACAGAAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTAC 56373
QY 139 AlaLysLeuLysGluGlnLeu 145
|||
Db 56372 TCAAAACCTTAGGTAGAAAT 56352
```

## RESULT 5

US-10-425-114-19518

; Sequence 19518, Application US/10425114

; Publication No. US2004003488A1

## ; GENERAL INFORMATION:

```
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 19518
; LENGTH: 1691
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-041-D8_FLI
US-10-425-114-19518
```

## Alignment Scores:

Pred. No.:	2,49e-22	Length:	1691
Score:	380.50	Matches:	110
Percent Similarity:	48.4%	Conservative:	106
Best Local Similarity:	24.7%	Mismatches:	155
Query Match:	12.0%	Indels:	75
DB:	8	Gaps:	14

US-10-644-084-2 (1-615) x US-10-425-114-19518 (1-1691)

```
QY 27 LysMetSerProSerSerLeuTyr-----SerGlnGlnValLeuCysSer 41
|||
Db 160 CGGATGTCCTCGCGCGGTTTCGATCTCCGGGCTCGTCAGCAGTTG-----CAG 213
QY 42 SerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGluAsn 61
|||
Db 214 CCGCCACCGTCCGACGCCGATGAGCAGCGGGGCGCTTCGCAACCGCGAGAAC 273
QY 62 IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrPheGlyPheProSerLeu 81
|||
Db 274 CTGGAGCACTGCGCCCGGTACTCTGAACGACGCTCGTCTTCGGATTCGCGCA--- 330
QY 82 TyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsn 101
|||
Db 331 -----TCTCTGACCTCTCGCACCGATCCGGTCTCGATTGCAAGAACTTGTAA 381
QY 102 CysMetAsnGluLeuValLeuValLeuValLeuValLeuValLeuValLeuValGlu 121
|||
Db 382 TGCATCTATGCGCTCTCGACGACGACAAAGGACATTTGAGTTTCAGGAGTCTACAAAT 441
QY 122 ThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeu 141
|||
Db 442 GACCAAGCGCAGCGATGCATCTGTATCTCCGCTGGAAGCTAAGATTGAGAGATG 501
QY 142 LysGluGlnLeuGluThrSerArgGluMetIleGlyLeuGlnGluArgAspArgGln 161
|||
Db 502 GATGCTCAATTAGCAGCGAAGATCGAGAGCTAGCCACTCTTACTCGAACGAGGCCAAA 561
QY 162 LeuGlnCysLysAsnArgSerLeuHisGlnLeuLeuAsnGluLysAspGluValGln 181
|||
Db 562 AACACTGCTGCTGAAGCTCTCAGATTGATTAAGCTGCAACAGAAACGATGAATTTTCAG 621
QY 182 LysLeuGlnAsnIleAlaSerArg-----AlaThrGlnTyrAsnHisAspValLys 199
|||
Db 622 AAAATG-----GTTATAGGAATCAGCAAGTACCCACCACCAATTCATGAATGAAG 675
QY 200 ArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLys 219
|||
Db 676 AAGAAAGAGAAAGAAATACATAAAGTTGCAGGAAAGCTAAATCAGCTATTGATGAGAAG 735
QY 220 LysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGly 237
|||
```

Db 736 AAAAAGGAATCATCCCGTTCTGGAATGGAGATAAATGAACCTTGTTACAGAAAGAGACGA 795  
Qy 238 LysArgGlySerTrpArgThrAspLysThrGluAlaAArgAsnGluAspGluMetTyrLys 257  
Db 796 CAACCGGGAACATGGGAATGAAAAAG-----AACGCAATGATTATTACAAA 843  
Qy 258 IleLeuLeuAspTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeu 277  
Db 844 ATGATTGTTGATGCATATGAAGTGAAGAAACAAGAGCTGATCAAGAGATGCAGATTTA 903  
Qy 278 LysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLys 297  
Db 904 CGGCACACTGTTACGTTCAATGCAGATGGAGATGCGTGCATTTCTCAAT----- 951  
Qy 298 LysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIle---Glu 316  
Db 952 -----GCCCGGAATGGGTCAATCACAATCTACTGTTACTGATTAATGGAAGA 996  
Qy 317 AspAspSerGlyGluLeu-----SerArgAspSerValTrpGlyLeuSer 331  
Db 997 CAGGAGTCAGGGTCTCTCAGTCTCCGCTTGGTGGCAAGACGGATGTTTTCAGCTTGCCC 1056  
Qy 332 CysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLys 351  
Db 1057 TTTTCATATGCCAGAGATCAGATAGAAAGAGAGTTTTCGCACAAAATGACTTCAATCAAG 1116  
Qy 352 SerHisValGluLysLeuAspAsn-----GlnAlaSerLysValHisSerGluGlyLeu 369  
Db 1117 GCACGAATGACAACTTCAAGATGCTCAGAAAGGTGCAGAAAGTGACTTCTGAGGCAACT 1176  
Qy 370 AsnGluGluAspValIleSerArgGlnAspHisGluGlnGlnLeuThrGluLysLeuGluLeu 389  
Db 1177 GACCGGAG-----CTTGAGCTTGAAGCT 1200  
Qy 390 GluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuGlnGlnGlnLeu 409  
Db 1201 CAACCTGGTTGAAGCAAGAGCATCATCCAAAGAGCAGGCGCTGCATCATGTCCAAGCACTTC 1260  
Qy 410 AlaThr----- 411  
Db 1261 AAGTCTGTAAAGCAAGTGGCAGGAGGCATAGCGGTCTGGATGGCGAGCGTGAGCATCT 1320  
Qy 412 -----ThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyr 425  
Db 1321 CGCGAGGTGTAAAGCATGTCTGTAGCAAGGAT---ATATCTGTTCTGCAAGACCTCCTG 1377  
Qy 426 LeuLeuGluGluLysGlu 431  
Db 1378 ATTTTAGTAGTACCAG 1395

RESULT 6

US-10-424-599-27277  
; Sequence 27277, Application US/10424599  
; Publication NO. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yinua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(5323)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 27277  
; LENGTH: 1567  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_124632C.1  
US-10-424-599-27277

Alignment Scores:  
Pred. No.: 3, 03e-22 Length: 1567  
Score: 279.00 Matches: 94  
Percent Similarity: 52.8% Conservative: 95  
Best Local Similarity: 26.3% Mismatches: 143  
Query Match: 12.0% Indels: 26  
DB: 8 Gaps: 8  
US-10-644-084-2 (1-615) x US-10-424-599-27277 (1-1567)  
Qy 56 PheCysThrGlyGluAsnIleGlnSerIleSerTyrLeuAspGlnGlnLeuThrThr 75  
Db 176 TTCCCGCATCGGCAATTTGGAGCATTCGCGTAAAGTATCTCAACCAATCGCTCGTCACG 235  
Qy 76 PheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsn 95  
Db 236 TTTGGTTTCCCGGCT-----TCGTCGATCTCTTCGCAAAATGACCCCGCTTCA 283  
Qy 96 IleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeu 115  
Db 284 ATTCCGAGGACTTGCATTTGCACTTTACGCTTTCGTCAGACAGACAGCGCGACGTGGAG 343  
Qy 116 AlaGlnGluSerValGluThrGlnAsnLeuLysLysGlySerAspMetAspHisLeuGln 135  
Db 344 TTTAGAGAGTCTGCTTAATGATCAAAAGACAACGACTGTTGTCTGCACATTTTCGAGATTGGAG 403  
Qy 136 SerCysTyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeu 155  
Db 404 GCCAAAGTGGAGAGGCTTGAAGGTCACACTCAAGTCAAAGACAGGAGATAGCGACTATT 463  
Qy 156 GlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsn 175  
Db 464 ACTAGACGGAAGCTAAACACACTGCAGCTCTGAAGGCCCAATTTGAGAAGCTGCAACAG 523  
Qy 176 GluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg-----AlaThrGln 193  
Db 524 CAGAGGATGAATCCAGAGAATG-----GTTATTGTAATCAGCAAGTAAAGACTCAA 577  
Qy 194 TyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHis 213  
Db 578 CAATGTCATGATGATGAAGAAAGAAAGAAATACATAAAGTTTCGAGAGAGGCTAAAC 637  
Qy 214 GlnLeuValMet---AsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrVal 232  
Db 638 CAAGTGTGTTGAGAGAAAGAGAGCTTAGTTCAGGCATGGAGATGAATGAATTTGCTT 697  
Qy 233 GlyArgAlaAspGlyLysArgLysSerTrpArgThrAspLysThrGluAlaArgAsnGlu 252  
Db 698 CAGAAGGAAGGAGGCAACGTTGGAACATGGAATGGAAAGAGCT-----GAC 745  
Qy 253 AspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMet 272  
Db 746 AATGATTTTATAAAAGATTTGGATGCTTATGATCAAAAAATCAAGAACTGATGCA 805  
Qy 273 GluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeu 292  
Db 806 GAGAATGCTGATTTAAGAGACATTTAGATCAATGCAGGTGGATATCGTGATTTCTTA 865  
Qy 293 SerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIle 312  
Db 866 AATGCTCTCAATGAGTACCAAG-----CAATCTTTCACTGGCAATGAAAGAGTT 916  
Qy 313 SerAspIleGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCys 332  
Db 917 GAAAGTGATCTTCACAACTCTCCATTGGTTCGGAAAAATGGATGATATTGACCTTCCTTT 976  
Qy 333 AspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSer 352  
Db 977 CACATGGCCAGAGATCAATAGAAAGAGTCTTCGCAACAAAAATGGCTTCTATAAAGGAG 1036  
Qy 353 HisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGlu 372  
Db 1037 CGTATGGTTCAATTG-----CAAGATGCTCAAAAAGAA 1069

QY 373 AspValIleSerArgGlnAspHisGluGlnGlnGluThrGluLysLeuGluLeuGluLeuGlu 392  
 Db 1070 GCAGAGATTACTTCAGAGGCTACTGAGAGGAACTTGAG---CTTGAAGCTCAACTGTGC 1126  
 QY 393 ArgCysLeuMetIleLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 410  
 Db 1127 GAACGACGAGCATCATACAGGACGAGGATCATTAATGTCTTAACATCTTGCA 1180  
 RESULT 7  
 US-10-425-115-157154  
 ; Sequence 157154, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 157154  
 ; LENGTH: 1805  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_74904C.1  
 US-10-425-115-157154  
 Alignment Scores:  
 Pred. No.: 4,47e-22 Length: 1805  
 Score: 378.00 Matches: 103  
 Percent Similarity: 49.0% Conservative: 100  
 Best Local Similarity: 24.9% Mismatches: 143  
 Query Match: 11.9% Indels: 68  
 DB: 9 Gaps: 12  
 US-10-644-084-2 (1-615) x US-10-425-115-157154 (1-1805)  
 QY 54 GlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrIleuAspGlnGluLeu 73  
 Db 287 GCGCGCTTCGCAACGCGGAACTTGGAGCACTGCGCCGCTGCTGAAACAGAGCGCTC 346  
 QY 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93  
 Db 347 GTCACCTTCGGATTCCCGCA-----TCTCTCGACCTCTTCGCCACCGATCCG 394  
 QY 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArgLysAsn 113  
 Db 395 GTCTCGATTGCAAGAACTTGTAACTGCACTATGCGCTGCTGCAGCAGCGACAGAGGGAC 454  
 QY 114 LeuLeuAlaGlnGluSerValGlnThrGlnAsnLeuLysLeuLysLeuLysLeuLys 133  
 Db 455 ATTGAGTTCAGGAGTCTACAAATGACCGCGCGAGGATGCAATCTGATATCTCACGG 514  
 QY 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnGluThrSerArgArgGluMetIle 153  
 Db 515 CTGGAGCTTAAGATGAGAGATGGATGCTCAATTAGCAGCGAAGATGCAAGAGCTAGCC 574  
 QY 154 GlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu 173  
 Db 575 ACTCTTACTCGAAGCGAGGCCAAACACACTGTGCTCTCAAGTCTCAGATTGATAAGCTG 634  
 QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAlaSerArg-----Ala 191  
 Db 635 CAACAGAAACGCTGATGAATTTTCAGAAATG-----GTTATAGGAATCAGCAAGTACGC 688  
 QY 192 ThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeuLysGluArg 211  
 Db 689 ACCGACAAATTCATGAATGAAAGAAAGAAAGAAAGAAATACATAAAGTTGCGGAAAG 748

QY 212 LeuHisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMetAspValLeu 229  
 Db 749 CTAATTCAGGTATTGATGAGAGAAAGAAAGAAATCATCCGTTCTGGAATGGAGATAATG 808  
 QY 230 AsnTyrValGlyArgAlaAspGlyLysArgGlySerTyrArgThrAspLysThrGluAla 249  
 Db 809 AACTTGTGTACAGAAAGAGGACGACAAACCGGAACATGGAATGGAAGAAAG----- 859  
 QY 250 ArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGln 269  
 Db 860 ---AACGACAATGATTATTACAAATGATTGTTGATGTCATATGACTGAAGAAACAAGAG 916  
 QY 270 IleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIle 289  
 Db 917 CTGATGCAAGAGATGCAATTTACGGGCACCTGTTACGTTCAATGCAATGCAATGCGT 976  
 QY 290 SerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309  
 Db 977 GACTTTCTCAAT-----GCCCGAATGGTCATCACA 1009  
 QY 310 ValAlaIleSerAspIle---GluAspAspSerGlyGluLeu-----Ser 323  
 Db 1010 TCTACTGTTACTGATAATGGAAGACAGGAGTCAGGCTCTCCTCAGTCTCGCTTGGTGGC 1069  
 QY 324 ArgAspSerValTyrGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIle 343  
 Db 1070 AAGACGGATGTTTTCAGCTTGCCTTTTCATATGGCCAGAGATCAGATAGAGAGATTG 1129  
 QY 344 ArgLysGlnTyrArgIleLeuLysSerHisValGluLysLeuAspAsn-----GlnAla 361  
 Db 1130 CGCAAAATGACTTCAATCAAGCGACGAATGACACAACTTCAAGATGCTCAGAAAGGT 1189  
 QY 362 SerLysValHisSerGlyGluLeuAsnGluGluAspValIleSerArgGlnAspHisGlu 381  
 Db 1190 GCAGAGTGAAGTCTCTGAGCAACTCACCAGGAG-----1222  
 QY 382 GlnGluThrGluLysLeuGluLeuLysGluArgCysLysGluMetIleLysAlaGln 401  
 Db 1223 -----CTTGAGCTTGAAGCTCAACTGTTGAAGCAAGAGCATCATCAAGAGCAG 1273  
 QY 402 GlnGlnLeuGlnGlnGlnLeuAlaThr-----411  
 Db 1274 GCCTGCATCATGCTCCAAGCACTTCAAGTCTGTATAAGCAAGTGGCAGGAGCATACGGT 1333  
 QY 412 -----ThrCysAspAspThr 417  
 Db 1334 CTGGATGGCGCGTGGAGCATCTCCCGAGGTGTAAAGCATGTCTGTAGCAAGGAT--- 1390  
 QY 418 ThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLysGlu 431  
 Db 1391 ATATCTGTTCTGCAAGACCTCCTGATTTTAGTAGTACCGAG 1432

## RESULT 8

US-10-437-963-34854/c  
 ; Sequence 34854, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: 38-21(53221)B  
 ; CURRENT APPLICATION NUMBER: US/10/437,963  
 ; CURRENT FILING DATE: 2003-05-14  
 ; NUMBER OF SEQ ID NOS: 204966  
 ; SEQ ID NO 34854

; LENGTH: 3750  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_3882C.1  
US-10-437-963-34854

Alignment Scores:  
Pred. No.: 3 17e-20 Length: 3750  
Score: 361.50 Matches: 96  
Percent Similarity: 51.8% Conservative: 92  
Best Local Similarity: 26.4% Mismatches: 144  
Query Match: 11.4% Indels: 31  
DB: 8 Gaps: 10

US-10-644-084-2 (1-615) x US-10-437-963-34854 (1-3750)

```
Qy 51 GlyValPheGlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAsp 70
Db 3561 GCGCGCGCGCGGCTTCGCGAGCGCGGGAACCTCGAGCACTGCGCCAGGTACCTCAAC 3502
Qy 71 GlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerIleSerLysGluAla 90
Db 3501 CAGACGCTGTCACCTTCGGCTTCGCGCC-----TCGCTCGATCTCTTCGCC 3454
Qy 91 LysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGln 110
Db 3453 ACGGACCCGATGATTCGAGACGCGTCACTGATGATGCTCTCTTCAGCAGCGG 3394
Qy 111 ArgLysAsnLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAsp 130
Db 3393 CAGCGTGATATTGAGTTCGCGAGTCTACCAATGATCTCGCTAGGTATGCGTCTGAT 3334
Qy 131 MetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuThrSerArgArg 150
Db 3333 ATTTCTAGATTAGACCAAGATTGAAGAATGGATGCTCAATTAGCTGCAAAAGATCGC 3274
Qy 151 GluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSerLeuHis 170
Db 3273 GAGTGGCCACATTGACTAGACGAGGCGCAAAACATGCACTTTGAGGCTCAGATT 3214
Qy 171 GlnLeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIleAlaSerArg 190
Db 3213 GACAAGCTCAACAGACGCGATGAGTTCCAAAAAATG-----GTTATTGGCAATCAG 3160
Qy 191 -----AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeu 208
Db 3159 CAAGTACGTACCCAGCAAAATTCATGAATGAAGAAAAAGAGAAAAATACATCAAAATTG 3100
Qy 209 LysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsn-----IleAlaMet 226
Db 3099 CAGGAGAGTTAAACAGGATTAATGGAGAAAGAGGAATCATCAGTTCCAGAAATG 3040
Qy 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 3039 GAAATAATCACTTGTTCAGAAAGAGGACGCGTGGAACTTTGGAGTGGAAAAAAG 2980
Qy 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 2979 -----AATGACAATGATTATTACAAATGATTTGATGCTCATGAGGTAAAG 2932
Qy 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLys 286
Db 2931 AGCAGAGTTGATGCAAGAAATGCTGATTTGCGAGCACTATTGCTTCGATGAGATG 2872
Qy 287 GluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGly 306
Db 2871 GATATGCGCGAGTTCCTCAATGCTCCAAATGGAGTTCCA-----CAGCCAGCTGTGTGGGC 2815
Qy 307 ThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSer 326
Db 2814 AATGGACGACGAGGACGAGTCTCTCTCAGTCTCCACTTTGGT-----GGCAAGACGGAT 2761
```

```
Qy 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346
Db 2760 GTCTTTGATTGCCCTTTTCACATGCCTAGACAGCAGATTGAAGAGAGTTTGGCCTAAA 2701
Qy 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsn-----GlnAlaSerLysVal 364
Db 2700 ATGGCTTCCATAAAGGCCGAATGACACAACTTCAAGATGCCCAAAGGGTGCAGAAAGTG 2641
Qy 365 HisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGluThr 384
Db 2640 ACATCAGAAAGTACC-----GAGCGGAATCTT 2614
Qy 385 GluLysLeuGluLeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeu 404
Db 2613 GAA---CTTGAAGCTCAACTAGTTGAAGCTAGACATTATCCAGGACGAGCATCCCTC 2557
Qy 405 LeuGlnGln 407
Db 2556 ATGTCCAAG 2548
RESULT 9
US-10-767-795-5235
; Sequence 5235, Application US/10767795
; Publication No. US20040181830A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 5235
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09MAY01-C71981_1
US-10-767-795-5235
Alignment Scores:
Pred. No.: 1 24e-18 Length: 1062
Score: 334.50 Matches: 95
Percent Similarity: 50.8% Conservative: 95
Best Local Similarity: 25.4% Mismatches: 140
Query Match: 10.6% Indels: 44
DB: 9 Gaps: 11
US-10-644-084-2 (1-615) x US-10-767-795-5235 (1-1062)
Qy 20 SerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeu 39
Db 3 TCCTCGCTCTCTCTCTCAGACAAA-----TACACACAGTTTCTCTCA 44
Qy 40 CysSer-----SerValProLeuSerLysAsnVal 49
Db 45 ATCTCTTCGCAATGCCGGCAACTGATGCCAATTTGGATCTCCCTCTCTCAGTCCGCC 104
Qy 50 HisGlyValPheGly---ValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyr 68
Db 105 TTCTCTATCGGAGAGTATACATTTGCAGATGCTGGAACTTTGGACCACTGTACAAAGTAC 164
Qy 69 LeuAspGlnGluLeuThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLys 88
Db 165 TTGAACACAGACGCTCGTTACCTTTGGCTTCTCTGCT-----TCTCTCGATCTC 212
Qy 89 GluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuVal 108
Db 213 TTTGCCAATGATCGGTTTCGTCGCGAGGACCTGCAATTCATATATATTCATTGCTGCAG 272
Qy 109 LeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGly 128
```



```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009265.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN_HIT: BE694886.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT_HIT: g114723835, EVALUE 3.00e-98
; OTHER INFORMATION: SWISSPROT HIT: Q40554, EVALUE 1.20e+00
US-10-029-386-5594

Alignment Scores:
Pred. No.: 1.69e-18 Length: 513
Score: 328.00 Matches: 76
Percent Similarity: 62.5% Conservative: 4
Best Local Similarity: 59.4% Mismatches: 13
Query Match: 10.4% Indels: 35
DB: 2 Gaps: 2

US-10-644-084-2 (1-615) x US-10-029-386-5594 (1-513)
QY 237 GlyLysArgGlySerTrpArgThrAspLysThrGluAla----- 249
Db 511 GGAAGAGAGGGCTCTCTGGAGGACTGGTAATACTGAAGCCAGGTAAGTCCAGTTTAAATT 452
QY 250 ----- 250
Db 451 GATAAATGTAATCAGACTTCTTCTATCTCTGCAATTTTCATCTGACAGTTTCTTCTCTCAG 392
QY 250 gGnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270
Db 391 GAATGAAGATGAAATGTATATAAATTTCTTCTGAATGATTATGAATATCGTCAGAAACAAAT 332
QY 270 eLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMetIleSe 290
Db 331 CCTAATGGAATGCAGAACTTAAGAGGTTCTTCAACAAATGAAAAAGAAATGATTTTC 272
QY 290 rLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrVa 310
Db 271 TCTTCTTCTTCTCCCAAGAAACCTAGAGAAAGAGTAGATGATGATGATGATGATGATGATGAT 212
QY 310 lAlaIleSerAspIleGluAspSerGlyGluLeu----- 322
Db 211 AAGTGGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 152
QY 323 -SerArgAspSerValTrpGly 329
Db 151 TGGTAGGCATTTCTCAATGGGA 130

RESULT 11
US-10-425-114-10108
; Sequence 10108, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 10108
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700894116_FLI
US-10-425-114-10108

273 CACGGCAGCCGACATTTAGAGAGTCTGCTTAATGAGCAAGACACACATTATTA 332
129 SerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSer 148
333 TCAGATATCTCAAGATTAGAGGCAAAAGTTGAGAGGGCTTGATGCACAGTTGCAAGCTAAA 392
149 ArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArgSer 168
393 GATAGAGAGATGACCAACAATGACTAGAACGGAAGCAAAAATACTCGAGCTTTTAAGGCA 452
169 LeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAla 188
453 CAGATTGAGAGAGTACAGCAAGAACGAGATGAATTTTCAGAGGATG-----GTGATTGGT 506
189 SerArg-----AlaThrGlnTyrAsnHisAspValLysArgLysGluArgGlnTyrAsn 206
507 AATCAGCAAGTGAAGACTCAACAGATACATGAATGAAAGAAAGCAAAAGGAGGTACATA 566
207 LysLeuLysGluArgGluHisGlnLeuValMet---AsnLysLysAspLysAsnIleAla 225
567 AAGTTCCAGAGAGGCTGAACCAAGTCTTAATGGAAGAAAGAAAGAAATCAAGATCAGGA 626
226 MetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAsp 245
627 ATGGAATATGAAATTTGCTTCGAAAGAAAGAGCGGCAACGTGGACCTGGAATGGGAAG 686
246 LysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyr 265
687 AAAGCA-----GACAATGATTTCTATAAAAGATTTGGGATGCTTATGAGGCA 734
266 ArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLys 285
735 AAAAATCAAGAACTGAAGACAGAAATGCTGATTTAAGAGCAATTAAGGCTCAATGCAG 794
286 LysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAsp 305
795 GTAGATATCGGTGATTTCTTAATGCTCTCTTAATGGTTGCCAAGCA----- 842
306 GlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSer----- 323
843 -----TCTTTGCTCTGTAGCGAAAGACATGAAATGATTCATCACAATCGCCACTT 893
324 -----ArgAsp-SerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAs 341
894 GTGGAGGAGCGGATGTCTTTCATCTCTCTTTTCATATGGCTAGGATCAATAGAGCA 953
341 nSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsn----- 359
954 AAGTCTTCCACTAGATGGCTTCTGTAAAGAGCGGATGGTTCAATTACAGGATCGCGCA 1013
360 -GlnAlaSerLysValHisSerGlyGlyLeuAsnGluGlu 372
1014 AAAAGGTGCTGAAGTCACCTCTCTGAGCAACACAGAGGGAG 1053

RESULT 10
US-10-029-386-5594/c
; Sequence 5594, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5594
; LENGTH: 513
; TYPE: DNA
```

## Alignment Scores:

Pred. No.: 1,01e-14 Length: 1199  
Score: 290.00 Matches: 72  
Percent Similarity: 53.5% Conservative: 82  
Best Local Similarity: 25.0% Mismatches: 112  
Query Match: 9.2% Indels: 22  
DB: 8 Gaps: 7

US-10-644-084-2 (1-615) x US-10-425-114-10108 (1-1199)

```
QY 126 LysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeu 145
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 8 CGACTGTTGCTGCATTTTCGAGATTGGAGGCCAAAGTCGAGAGCGCTTGAAGTCAACTA 67

QY 146 GluThrSerArgGluMetLeuGlyLeuGlnGluArgAspArgGlnLeuGlnCysLys 165
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 68 CAAGTCAAAAGACAGGAGAGATGCGACTATTACTAGAACGGAAGCTAAACACACTGCAGCT 127

QY 166 AsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsn 185
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 128 CTGAAGCCCAATTGGAAGTGCACACAGAGAGGATGAATTCAGAGAATG----- 181

QY 186 IleIleAlaSerArg-----AlaThrGlnTyrAsnHisAspValLysArgLysGluArg 203
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 182 GTTATTGGTAAATCAGCAAGTAAGACTCAACAAATGCATGATGAAGAAGAGAAAAA 241

QY 204 GluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMet---AsnLysLysAspLys 222
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 242 GAATACATAAAGTTGCAGGAGAGGCTAAACCAAGTGTTCATGAGAGAAGAAAGATCT 301

QY 223 AsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrp 242
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 302 AGGTCAGGATGAGATAATGAATTTGCTTCAGAGGAAGGAGGCAAGCTGAACATGG 361

QY 243 ArgThrAspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAsp 262
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 362 AATGGGAAGAAAGCT-----GACAATGATTTTATAAAAGATTCGTGGATGCT 409

QY 263 TyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGln 282
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 410 TATGAGTCAAAAATCAAGAACTGATGGCAGAGAATGCTGATTTAAGACATATTATTGAGA 469

QY 283 GlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArg 302
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 470 TCAATGCAGTGGATATGCTGATTTCTTAATATGCTCCTTAATGAGTACCAAG----- 523

QY 303 AlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeu 322
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 524 ---CAATCTTCACTGGCAATGAAGAGTTGAAAGTGATCCTTCACAATCTCCATTGGTT 580

QY 323 SerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSer 342
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 581 GGGAAATCGATGTATTTGACCTTCCTTTTCATGCGCAGAGATCAATATAGAAGAAAGT 640

QY 343 IleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSer 362
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 641 CTTCCGCAACAAATGCGCTCTATAAAGGAGCGTATGGTTCAATG----- 685

QY 363 LysValHisSerGluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGln 382
      ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 686 -----CAAGATCTCAAAAAGAAAGCAGAAGTTACTTCAGAGCGCTACTGAGAGG 733

QY 383 GluThrGluLysLeuGluLeuGluLleGluArgCysLysGluMetIleLysAlaGlnGln 402
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 734 GAACATTGAG---CTTGAAGCTCAACTTGTGGAAGCAAGGAGCATCATACAGGAGCGCA 790

QY 403 GlnLeuLeuGlnGlnLeuAla 410
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 791 TCAATATGCTTAACATCTTGCA 814
```

RESULT 12

US-10-029-386-19350/c

; Sequence 19350, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 19350  
; LENGTH: 178  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC009265.5  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
; OTHER INFORMATION: EST\_HUMAN HIT: AU133150.1, EVALUE 1.00e-95  
; OTHER INFORMATION: SWISSPROT HIT: P03654, EVALUE 1.10e+00  
; OTHER INFORMATION: NT HIT: gi14723825, EVALUE 8.00e-96  
US-10-029-386-19350

## Alignment Scores:

Pred. No.: 2.09e-15 Length: 178  
Score: 285.00 Matches: 56  
Percent Similarity: 96.6% Conservative: 1  
Best Local Similarity: 94.9% Mismatches: 2  
Query Match: 9.0% Indels: 0  
DB: 7 Gaps: 0

US-10-644-084-2 (1-615) x US-10-029-386-19350 (1-178)

```
QY 251 AsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIle 270
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 177 AATGAAGATGAATGTATAAATTTCTTGAATGATGATATCGTCAGAACAAATC 118

QY 271 LeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGluMetIleSer 290
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 117 CTAATGGAATATCGAAGCTTAAAGAGGTTCTTCAACAAATGAAAAAGGAAATGATTTCT 58

QY 291 LeuLeuSerProGlnLysLysProArgGluArgAlaGluAspGlyThrGlyThr 309
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
Db 57 CTTCTTTCTCCCCAAAGAAACCTAGAGAAAGATAGATGATACAGGAAC 1
```

## RESULT 13

US-10-369-493-45767  
; Sequence 45767, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 45767  
; LENGTH: 5040  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-45767

## Alignment Scores:

Pred. No.:	4.28e-09	Length:	5040
Score:	234.50	Matches:	145
Percent Similarity:	36.4%	Conservative:	124
Best Local Similarity:	19.6%	Mismatches:	230
Query Match:	7.4%	Indels:	241
DB:	7	Gaps:	28

  

US-10-644-084-2 (1-615) x US-10-369-493-45767 (1-5040)			
QY	15	GluAsnLysAsnLeuSerGlnTyrThr	23
DB	3028	GAATAATAAGATCTATCTGAATTTGGTTATCCGCTTGAAAGGATCGGCTGATGCCAA	3087
QY	24	SerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCysSerVal	43
DB	3088	GCAGAATTAAACAAAAACCAATCTCTATTGTATAGCGCTCAGGATTTA	3135
QY	44	ProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu	60
DB	3136	---CTGATTAACATGAAGGAATGGATGGAGAAAGAGCTGATTATGAAGAGAAATTG	3192
QY	61	---AsnIleGluGlnSer	68
DB	3193	ATATCTAATATTGAACAGCGGAATCTTCGCGTGTGAAATTCGGTATTGATTGAAAG	3252
QY	69	LeuAspGlnGluLeuThrPheGly	78
DB	3253	GTTGATGACACAGCAGCAATAATATGCGATGAAGATCAATTTGAAGCTAGTTCTTTGGTTT	3312
QY	79	ProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla	98
DB	3313	TCAACTTACGTCATCAAGAAATCTTTAGAGACGAAG	3351
QY	99	ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGlu	118
DB	3352	TTAACCACCTGTGAAGAGAGATTTGGCATTCGTTAAACAAAAGAACGATAGTTAGAAAAA	3411
QY	119	SerVal---GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCys	137
DB	3412	ACGATAAATGACCTGCAACGAACCCAGACTTTATCGGAGAAGGAATAC	3465
QY	138	TyrAlaLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGly	154
DB	3466	TCGCGAGTTATTATTCACGAATTCAGAGGATATCACGAAGAAGTTACTCAAGTTACATTT	3525
QY	155	-----LeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArg	167
DB	3526	TTAANAAGAGANAATGCCATCTTACAAAGAGTCTCAAAAACGTAAACCGAANAACACAGG	3585
QY	168	SerLeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIle	187
DB	3586	GAATAATTATAACAATTAATGATAGCAAGAGAAATTTCTCGATTACAGCGTGATCTC	3645
QY	188	AlaSerArgAlaThrGlnTyr	204
DB	3646	ATACAAACAAAAGAACAGTTTCGATAAATTTCAACAAAATTTTACTATATGAAGTGAA	3705
QY	205	TyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIle	224
DB	3706	ATGGACCAATGTAAACAGAGGTACCGAGCTTATCTCAACACAAAAGAT	3756
QY	225	AlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThr	244
DB	3757	-----GCCCAAAAAAAA	3768
QY	245	AspLysThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGlu	264
DB	3769	GACATTTGAAGAAGCTAACTAATGAG	3807
QY	265	TyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMet	284
DB	3808	GGAAAACTTTCAGCGCTGAGAACGCAACGCTTGACTTGGAAAAAATTCATAGGCTA	3867

QY	285	LysLysGluMetIleSerLeuLeuSerProGlnLysLysLysLysProArgGluArgAlaGlu	304
DB	3868	AAGAAGCAGCGCTCATGAAAAATTTGGACGCATCTTAAGAACAA	3909
QY	305	AspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGlyGluLeuSerArg	324
DB	3909	-----	3909
QY	325	AspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArg	344
DB	3910	-----CRAAGCTGCTTTAACTAACGAACCTCAAC	3936
QY	345	LysGlnTrpArgIleLeuLysSerHisValGluLysLysLeuAspAsnGlnAla	361
DB	3937	GAA-----TTAAAGCAATTTAAAGACAAATTAGAACCAAGATCTACACTTTGAA	3984
QY	362	-----SerLysValHisSerGluGlyLeuAsnGluGluAsp	373
DB	3985	AACGCTAAAGTTATTGATCTAGATCGAAGCTCAAGCACATGAACCTCAATCAGAAGAC	4044
QY	374	ValIleSerArgGlnAspHisGluGlnGluThr	384
DB	4045	GTC-----TCGAGAGATCACGAAAAGGACACTTATCGGCACATTAATGGAGGAAATTGAA	4098
QY	385	-----	Glu 385
DB	4099	TCACTCAAAAAGGAGTTACAAAATTTTCAAAACTGCGAATAGCTCATCTGATGCTTTGAG	4158
QY	386	LysLeuGluLeuGluIleGluArgCysLysGlu---MetIleLysAlaGlnGlnGlnLeu	404
DB	4159	AAATTTGAAATTAACATGGAGAAAGAAAAGGACAGAAATTTATTGACGAAAGACGAAGAA	4218
QY	405	LeuGlnGlnLeuAlaThrCysAspAspAspThrThrSerLeuLeuArgAspCys	424
DB	4219	TTTGAAGAAAAATTTACAGAGACCTTAATTAAGAGTACAAGCTCA	4269
QY	425	TyrLeuLeuGluGluLysGluArgLysGluGluTrp	437
DB	4270	GAATACAGTAAAGATATCGAAACATTTGMAAAAGGAATGCTAAAGGAATATGAAGATGA	4329
QY	438	ThrLeu-----PheLysGlnLysLysAsnPheGluArgGluArgSerPheThr	455
DB	4330	ACGTTAAGGAGAAATTAAGGACAGAGAGAAATTTAAAGAAAGAAATAGAGTTACCAAGC	4389
QY	456	GluAlaIleAlaIleArg-----LeuGlyLeuGluArgLysAlaPheGluGluArgAla	473
DB	4390	GAGGAAAGGATACAAAGATCATTAGTAGCGCTAAAGAGAATTAGAGAAGAAATTCGA	4449
QY	474	SerTrpValLysGlnGln	479
DB	4450	AAAGAAATTTGAAAGAAAAACGCGGGAAGCTTAAACGTTCTTTGATAACAAAGTAGCGGGAA	4509
QY	479	-----	479
DB	4510	GATCGCGAAGAGAATATTGGAACAGTCCATCGAAGGGGAATTTCTGAGAGCGCGTCCGCT	4569
QY	480	-----PheLeuAsnMetThrAsnPheAspHisGlnAsnSer---GluAsnValLys	495
DB	4570	GTCGCTGGCTTTTATAATCAAAAAACCTTAAACCTCAAGACAAATTTGAAGATGTGAAG	4629
QY	496	LeuPheSerAlaPheSerGlySer-----SerAspProAspAsnLeuIle	510
DB	4630	AATGACGTTTCTTCAATGACAGCCAGTCAATGGTTTACAAATAAAGAAAAATAATATCGTG	4689
QY	511	ValHisSerArgProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCys	530
DB	4690	GATAGTAGCGCTGCTGGTAACAGGCCATTTCCAACTTTAGTTTGGTGAAGCCATCTTC	4749
QY	531	ThrSerLysLeuThrLysSerLeuProAlaSerProSerThrSerAspPheArgGlnThr	550
DB	4750	TCCTCTAAT---ACATCATCTCTTCAAGT-----TTTCAGAACCCATTCACGATCC	4800



```
Db 661 AAAGAGAGGAATAGCCTCCAGAGGAATCGTTGAAACCATATGAGAAAAATTTTGCAT 720
QY 176 GluLysAspGluVal
Db 721 GAAAAGGAGGAGGATATTACTAAAGGAACAAGCTCTACTCAATCAGAGGGACGAGAACATC 780
QY 187 IleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsn 206
Db 781 CTT---GAGAGATTAGCTTATGTTACTCTCAGAGAAAAGAGTGGAGAGAAAAGAAT 837
QY 207 LysLeuLysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIleAlaMet 226
Db 838 ATTCCTTGAAGCTGAACGGAAGGTTTGTAGAGGAAAAATATAAGTTGGAGCTGAANAATG 897
QY 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 898 GAGGCAATA----- 906
QY 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 907 ---GTTTCAAGAGAGGAGCTCTGATTCAGAGGAATCCCTGCTTGAC-----AAAAGG 957
QY 267 GlnLysGlnIleLeuMet
Db 958 GAGAGTGAATTAATTTTGCAGAGACAAATCGCAAGTAAAGAACGGCCGCGAGATTGAA 1017
QY 279 LysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLys 298
Db 1018 AGGTTGAATCAGGAA-----CAAGCTATAGCCTTG----- 1047
QY 299 ProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAsp 318
Db 1048 ---GAGAGGAGAAAAATGATTTTGAATCTGAGATCGCAACAAACAAATGCTTTTGAT 1104
QY 319 SerGly---GluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGlu 337
Db 1105 GCAGCAATGGAGGTGACAGAAATGCGTGCATCAGAGAGATGT---GCTCTCAGTGAG 1161
QY 338 GlnLeuThrAsnSerIleArgLysGlnTrpArgIle----- 349
Db 1162 CAGGAATCTGTGGTTGTGCAAGAGATCCAGAACTCTTGACCTTCAGCTAGCTGAATTAGCA 1221
QY 350 -----LeuLysSerHisValGluLysLeu 357
Db 1222 AGCAAGAAAGCGGTTGCGAGGAGATCAGATGAGTTGAAAGAGAGGAGGAGGAAAAAGCTC 1281
QY 358 AspAsnGlnAlaSerLysValHisSerGluGlyLeuAsnGluGluAspValIleSerArg 377
Db 1282 TTGTTACACAGAGAACCCATACAAATGAACTTCAAAAAGAAAGGGAAGAAATACAAAGG 1341
QY 378 GlnAspHisGluGlnGluThrGluLysLeuGluIleGluArgCysLysGluMet 397
Db 1342 ATTAATAATCGGATTGAGAGAGGAAAGAGGCTTCTTTGAGAGAGAGAAA---CGGGAAGCA 1398
QY 398 IleLysAlaGlnGlnLeu-----Leu 405
Db 1399 ATTCAGGCCCAACAAGATCTAGCAATAACACAGCAGATAGAGATGAGTTGCTTACTCTG 1458
QY 406 GlnGlnGlnLeuAlaThrThrCysAspAsp-----AspThrThrSerLeuLeuArg 422
Db 1459 CAGATGAACTTAAGAGAAATTTGACGCTCAGAGCCCAAAAAGGGAACCTCATGGCT 1518
QY 423 AspCysTyrLeuLeuGlu---GluLysGluArgLeuLysGluGluThrLeuPheLys 441
Db 1519 GATGCAGATAGGCTGCAAGCCGAAAGAGAAAGGTTTGAGATTGAATGGGAGCTGATTGAT 1578
QY 442 GluGlnLysLysAsnPheGluArg-----GluArgArgSerPhe 454
Db 1579 GAAAAGAAAGAGGAGCTACAAAAGGAAGCGCATCAAAATTTGCTGAAGAACGAAGCAATA 1638
QY 455 ThrGlu----- 456
```

```
Db 1639 ACTGAGTATCTGAGAAATGAATCTGATATCATCAACAGGAGAGGATAATCTCCGTGTT 1698
QY 457 -----AlaAlaIleArgLeuGlyLeuGluArgLysAlaPhe----- 468
Db 1699 CAGTTTCAAAAGTAATTCAGAAAACACTCTCTCGTGAACACAAAGAGTTTCATGAGTAAGATG 1758
QY 469 GluGluGluArgAlaSerTrp-----ValLysGlnGlnPheLeuAsnMet---Thr 484
Db 1759 CAGCAAGAACATGCAAGTTGGCTGAGTAAGATTCAACAAGAAAGCAAGATCTGAAAGAGA 1818
QY 485 AsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSer 504
Db 1819 GACATTGATATCCAGAGGGTGAA-----TTGCTAAATTCCTGCTAAGGCAAGCCAGATG 1872
QY 505 AspProAspAsnLeuIleValHisArgProArgGlnLysLysLeuHisSerValAla 524
Db 1873 GAAATAGATTCTTATTGT-----AGGGAAGGGAGGAAGAGTTTGAG----- 1914
QY 525 AsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSerThr 544
Db 1915 -----CAGAAAAAGGCCAAGGAACTCGAACACATCAATTTCTCAG 1953
QY 545 SerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSerValLeuAsn 564
Db 1954 AAGGAGATGATCAACACAAAA-----TTAGAACATGTTGCAGTTGATTCAGAGAAA 2004
QY 565 IleThrProGluSerLysProSerGluValAlaArgGluSerThrAspGln----- 582
Db 2005 CTT---AAGGATGAGAGAAAAGAGCTACTTTTGGAACTGTGAGAGGAGAGCAAGAGTTG 2061
QY 583 -----LysTrpSerValGlnSerArgProSerSerArgGluGlyCysTyrSerGly 599
Db 2062 TCTGAGATAAAGGCCACTATTGAAGCCTTTGAATTAATCAACGGGAG-----AAG 2109
QY 600 CysSerSerAlaPheArgSerAlaHisGlyAspArgAsp 612
Db 2110 CTGCAAGAGCAAAAGAAACTATTACATTCAGACCCGAGAA 2148
```

Search completed: June 14, 2006, 02:18:27  
Job time : 1992 secs

**THIS PAGE LEFT BLANK**

```

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: June 14, 2006, 00:53:30 ; Search time 238 Seconds
(without alignments)
912.379 Million cell updates/sec

Title: US-10-644-084-2
Perfect score: 3165
Sequence: 1 MGDWTVTPVLCTENKNLS.....CYSGCSSAFSAHGDRDDLDP 615

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 290227 seqs, 117694381 residues

Total number of hits satisfying chosen parameters: 580454

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlp
-Q=/abes/ABSSWEB/spool/US10644084/runat_12062006_150434_1613/app_query.fasta.1
-DB=Published Applications NA New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=ptco -NOR=ext -HEADSIZE=500 -MINLEN=50
-MAXLEN=2000000000 -HOST=abs8059p
-USER=US10644084 @CGN 1 1 29 @runat_12062006_150434_1613 -NCPU=6 -ICPU=3
-NO MAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*
1: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US09_NEW_PUB_seq.*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US06_NEW_PUB_seq.*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US07_NEW_PUB_seq.*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US08_NEW_PUB_seq.*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpna/PCT_NEW_PUB_seq.*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US10_NEW_PUB_seq.*
7: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US11_NEW_PUB_seq.*
8: /EMC_Celerra_SIDS3/ptodata/2/pubpna/US60_NEW_PUB_seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description
-----
1 359.5 11.4 1556 6 US-10-449-902-2430 Sequence 2430, Ap
2 356 11.2 1601 6 US-10-449-902-27762 Sequence 27762, A
3 349 11.0 1497 6 US-10-953-349-6159 Sequence 6159, Ap
4 231.5 7.3 5373 7 US-11-217-529-5516 Sequence 5516, Ap
5 230.5 7.3 5040 7 US-11-217-529-77055 Sequence 77055, A
6 215.5 6.8 1833 7 US-11-217-529-80544 Sequence 80544, A
7 209 6.6 2479 7 US-11-293-697-1966 Sequence 1966, Ap
8 207 6.5 3270 6 US-10-480-962-24 Sequence 24, Appl
9 206.5 6.5 10211 6 US-10-505-928-326 Sequence 326, Appl

```

```

10 205 6.5 5558 7 US-11-217-529-6026 Sequence 6026, Ap
11 201.5 6.4 5787 7 US-11-217-529-76923 Sequence 76923, A
12 192 6.1 4925 7 US-11-289-102-97 Sequence 97, Appl
13 190.5 6.0 5028 6 US-11-217-529-5947 Sequence 5947, Ap
14 188 5.9 1648 6 US-10-953-349-20044 Sequence 20044, A
15 187 5.9 2773 7 US-11-293-697-2130 Sequence 2130, Ap
16 185 5.8 9588 6 US-10-505-928-101 Sequence 101, App
17 183 5.8 3039 6 US-10-449-902-28253 Sequence 28253, A
18 181 5.7 1684 6 US-10-953-349-20040 Sequence 20040, A
19 181 5.7 2355 6 US-10-449-902-20065 Sequence 20065, A
20 178.5 5.6 1462 6 US-10-953-349-24038 Sequence 24038, A
21 178.5 5.6 2268 7 US-11-217-529-6178 Sequence 6178, Ap
22 178 5.6 2531 7 US-11-217-529-560 Sequence 560, App
23 177 5.6 4263 7 US-11-327-900-5 Sequence 5, Appl
24 176 5.6 3044 7 US-11-293-697-2049 Sequence 2049, Ap
25 175 5.5 2025 7 US-11-217-529-81827 Sequence 81827, A
26 175 5.5 4251 7 US-11-217-529-3111 Sequence 3111, Ap
27 174 5.5 4455 6 US-10-504-973-70 Sequence 70, Appl
28 173.5 5.5 1625 6 US-10-953-349-17058 Sequence 17058, A
29 173.5 5.5 3720 6 US-10-449-902-17102 Sequence 17102, A
30 173 5.5 1798 7 US-11-293-697-1637 Sequence 1637, Ap
31 172.5 5.5 1647 7 US-11-217-529-3824 Sequence 3824, Ap
32 171.5 5.4 2466 7 US-11-217-529-78674 Sequence 78674, A
33 170 5.4 1257 7 US-11-217-529-80717 Sequence 80717, A
34 169 5.3 1880 7 US-11-293-697-2304 Sequence 2304, Ap
35 169 5.3 3735 7 US-11-217-529-2189 Sequence 2189, Ap
36 168.5 5.3 2607 7 US-11-217-529-4263 Sequence 4263, Ap
37 168 5.3 2930 7 US-11-293-697-643 Sequence 643, App
38 167.5 5.3 1928 7 US-11-293-697-1901 Sequence 1901, Ap
39 167 5.3 2477 6 US-10-953-349-11093 Sequence 11093, A
40 165 5.2 1203 7 US-11-217-529-79694 Sequence 79694, A
41 165 5.2 2520 7 US-11-217-529-1961 Sequence 1961, Ap
42 164.5 5.2 2796 7 US-11-217-529-81089 Sequence 81089, A
43 164 5.2 2072 6 US-10-953-349-11127 Sequence 11127, A
44 163.5 5.2 1717 6 US-10-449-902-2456 Sequence 2456, Ap
45 163.5 5.2 3374 6 US-10-449-902-23271 Sequence 23271, A

```

## ALIGNMENTS

```

RESULT 1
US-10-449-902-2430
; Sequence 2430, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-Oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2430
; LENGTH: 1556
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK060264
; DATABASE ENTRY DATE: 2001-12-06
US-10-449-902-2430

Alignment Scores:
Pred. No.: 1.13e-23 Length: 1556
Score: 359.50 Matches: 105
Percent Similarity: 48.7% Conservative: 97
Best Local Similarity: 25.3% Mismatches: 163

```

```
Query Match: 11.4% Indels: 50
DB: 6 Gaps: 11
US-10-644-084-2 (1-615) x US-10-449-902-2430 (1-1556)

Qy 18 AsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSerGlnGln 37
Db 2 AATCAATCAATCTCTCTCTCCAGATCTCTCCACCCACCTCGCTCCCTCCATGTCGCG 61
Qy 38 ValLeuCysSerSerValPro-LeuSerLysAsnValHis----- 50
Db 62 CTCCTCGCGCTTCGATCTCTCGGCGCTCGTCGACCCGACGACGCGCATCGGAGCGG 121
Qy 51 -----GlyValPheGlyValPheCysThrG1 59
Db 122 AGCGGAGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
Qy 59 yGluAsnIleGluInSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPhePr 79
Db 182 GGGGAACCTCGAGCACTGCGCGAGGTACCTCAACAGAGCGCTCGTCACTTCGCTTCCC 241
Qy 79 oSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaVa 99
Db 242 GGGC-----TCGCTCGATCTCTTCGCCACCGACCGCTGATCGATTGCGAGAAC 289
Qy 99 lLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnLysE 119
Db 290 GTGCACTGATCTATGCTCTCTTCAGCAGCGGAGCGGATGATGATGATGATGATGATGATGAT 349
Qy 119 rValGluThrGlnAsnLeuLysLeuGlySerMetAspMetAspHisLeuGlnSerCysTyrAl 139
Db 350 TACCAATGATCTCGCTCAGCGTATGCGAGTCTGATATTTCTAGATTAGTAGCCAGATTGA 409
Qy 139 aLysLeuLysGluGlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnLysArgAs 159
Db 410 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469
Qy 159 pArgGlnLeuGlnCysLysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspG1 179
Db 470 GGGCAAAACACATGCACTTTGAGAGCTCAGATTGCAAGCTGCAACAGGAGGAGGATGA 529
Qy 179 uValGlnLysLeuGlnAsnIleAlaSerArg-----AlaThrGlnTyrAsnHisAs 197
Db 530 GTTCCAAAAATG-----GTTATTGGCAATCAGCAAGTACGTACCCAGCAAAATTCATGA 583
Qy 197 pValLysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMe 217
Db 584 AATGAAGAAAAAGAGAAAGAAATACATCAAAATTCAGGAGAGAGTTAAACCCAGGTATTAAT 643
Qy 217 tAsnLysLysAspLysAsn-----IleAlaMetAspValLeuAsnTyrValGlyArgAl 235
Db 644 GGAGAAAAAGAGGAATCATCATCGTTCAGGATGGAATGGAATGGAATGGAATGGAATGGAATGGA 703
Qy 235 aAspGlyLysArgGlySerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMe 255
Db 704 AGGACGCGAGCGTGAACCTTGAGTGGAAAAAAG-----AATGACAATGATTA 751
Qy 255 tTyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAl 275
Db 752 TTACAAAATGATTTGATGCTTATGAGGTAAAGAAAGCAAGAGTTGATGCAAGAAATGTC 811
Qy 275 aGluLeuLysLysValLeuGlnMetLysLysGluMetIleSerLeuLeuSerProG1 295
Db 812 TGATTTGCGAGCACTATTGCGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 871
Qy 295 nLysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspI1 315
Db 872 AAATGGAGTTCCA---CAGCCAGCTGTTGTGGCAATGGACACAGGAGGAGGATCTCC 928
Qy 315 eGluAspAspSerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrVa 335
Db 929 TCAGTCTCCACTTGGT-----GGCAAGACCGATGCTTTGATTTGCGCTTTTCACATGGC 982
```

```
Qy 335 lArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValG1 355
Db 983 TAGAGACCAGATTGAAGAGAGTTTTCGCGCACTTAAATGCTTCCATAAAGGCCGAATGAC 1042
Qy 355 uLysLeuAspAsn-----GlnAlaSerLysValHisSerGluGlyLeuAsnGluLys 373
Db 1043 ACAACTTCAAGATGCCAAAGGGTGCAGAAAGTGCATCAGAAAGTACC----- 1091
Qy 373 pValIleSerArgGlnAspHisGlnGlnThrGluLysLeuLeuGluLysLeuGluLys 393
Db 1092 -----GAGCGGGAACCTTGA---CTTGAAGCTCAACTAGTATGTA 1126
Qy 393 gCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGln 407
Db 1127 AGCTAGAGCATTTATCCAGGAGCAGGCATCCCTCATGTCCAAG 1169

RESULT 2
US-10-449-902-27762
; Sequence 27762, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agricultural Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205V1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: JP 2002-203269
; PRIOR FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 27762
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Oryza sativa
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AK103204
; DATABASE ENTRY DATE: 2002-08-28
US-10-449-902-27762

Alignment Scores:
Pred. No.: 2,43e-23 Length: 1601
Score: 356.00 Matches: 103
Percent Similarity: 49.9% Conservative: 96
Best Local Similarity: 25.8% Mismatches: 154
Query Match: 11.2% Indels: 46
DB: 6 Gaps: 12

US-10-644-084-2 (1-615) x US-10-449-902-27762 (1-1601)

Qy 28 MetSerProSerSerLeuTyrSerGlnGlnValLeuCysSerSerValPro----- 44
Db 51 ATGTCGCGCTCTCTCGCGCTTC-----GATCTCGCGGCTCTGTCGACCCGACCCAGCAC 104
Qy 45 -----LeuSerLysAsnValHisGlyValPheGly 54
Db 105 GCCATCGGAGCGGAGCGGAGCGGAGCGGATGACGTGCGGTGGCGCGCGCGCG 164
Qy 55 ValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThr 74
Db 165 CGGTTCGCGGAGCGGCGGGAACCTCGAGCACTGCGCAGGATGCTCAACAGAGCGTCTGTC 224
Qy 75 ThrPheGlyPheProSerLeuTyrGluGlnSerLysSerLysGluAlaLysArgGluLeu 94
Db 225 ACCTTCGGCTTCCCGGCC-----TCGCTCGATCTCTTCGCCACCCAGCCCGGTA 272
Qy 95 AsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeu 114
Db 273 TCGATTGCGAGACGTCGCAACTGTCATCTCTCTTCTTCAGCAGCGGCGCGGTATATT 332
```





Qy	278	ysLysValLeuGlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysL	298
Db	839	GAGCGTTACTACGATCAACGACGAGGTGACATCGCTTCTTCTTGAAT----	885
Qy	298	ysProArgGluArgAlaGluAspGly-----ThrGlyThrV	310
Db	886	-----GCTTCTGGTGGATTACCAACCAATCTTTCGTAGCTAATGGAAGAC	931
Qy	310	alAlaIleSerAspIleGluAspAspSerGlyLeuLeuSerArgAspSerValTrpGlyL	330
Db	932	ACGGTCGACACCCCTTCTCAGTCTCCTTTGGGC-----GGGAAGACGACGCTTTTTCACC	985
Qy	330	euSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrpArgileL	350
Db	986	TACCTTTTCGTATGGCAGCGAGGTCAATAGAAGATAGTTTACGCTCTAAAGATGGTTTCCA	1045
Qy	350	euLysSerHisValcLulLysLeuAspAsnGlnAlaSerLysValHisSerGluGlyLeuA	370
Db	1046	TCAAGGAACGTTATGGGTCAAGTTACTAGATGCACAAAAAGAAGTATCC-----	1092
Qy	370	snGluGluAspValIleSerArgGlnAspHisGluGlnGlnLeuThrGluLysLeuGluLeuG	390
Db	1093	-----ATTACTTCAAGACGCTCAGAGAGGAAGCTTGAA---CTTGAAGCAC	1135
Qy	390	luIleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnLeuA	410
Db	1136	AACTCGTGAGGCACGCGACGATTTATCAAGNACAGGAATCTATTAATGTCCAACATTTC	1195
Qy	410	laThrThr	412
Db	1196	CTAAGACA	1203

## RESULT 4

```

US-11-217-529-5516
; Sequence 5516, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5516
; LENGTH: 5373
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-5516

```

Alignment Scores:		
Pred. No.:	2,449-11	Length:
Score:	231.50	Matches:
Percent Similarity:	38.6%	Conservative:
Best Local Similarity:	21.1%	Mismatches:
Query Match:	7.3%	Indels:
DB:	7	Gaps:
		33
		5373

US-10-644-084-2 (1-615) X US-11-217-529-5516 (1-5373)

Qy 15 GluAsnLysAsn-----LeuSerGlnTyrThrSerGluThrLysMetSerProSer 31  
|||:::|||||  
Db 3097 GAACACAAAACGAAGAACATCTCTCAAGTACAAATTCGAAAGGATGAGTATGAGTCA 3156  
|||:::|||||  
Qy 32 SerLeuTyrSerGlnGlnValLeuCysSerSerVal----- 43  
|||:::|

```
QY 308 GlyThrVal-----AlaIleSerAspIleGluAspSerGlyGluLeuSerArgAsp 325
Db 4231 AGACTAATTGATACAAGTAGAGTGCCTGGAGAAAGCCACT----- 4272
QY 326 SerValTrpGlyLeuSerCysAspThrValArgGluLeuThrAsnSerIleArgLys 345
Db 4273 -----TTGAGCCATGATGAATTTTGAAGAAGAAAGCGGATGAGATTATAAA 4320
QY 346 GlnTrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysVal--- 364
Db 4321 ATGAAAGACCAAAATCTCTCATCAAAAGAAAGATAAATCAAGCAAAACAGAACTACTA 4380
QY 365 -----HisSerGluGlyLeuAsn-----GluGluAspValIleSerArg 377
Db 4381 TCAACGGAAACAACCATGAAATGGGTAAATAGTTTGAAGGAACAACCTTGAGGCTATT 4440
QY 378 GlnAspHisGluGlnGluThrGluLysLeuGluGluIleGluArgCysLysGluMet 397
Db 4441 CAAGATCTAAATCCGAGGTTGAAGAGAAATTAAGAGACATGAGGAGAAATCTGCTCAC 4500
QY 398 IleLysAlaGlnGlnLeuLeuGlnGlnGlnLeuAlaThrThrCysAspAspThr 417
Db 4501 CTCAAATCAGATTTAAAGAAATCAAAAGAAACTGCGAAAGACTTAAAGAGTGACATTGAA 4560
QY 418 ThrSerLeuLeuArgAspCysThrLeuLeuGluGluLysGluArgLeuLysGluTrp 437
Db 4561 AACATGAAAGAAAGAAATAGAACTTTTAGTAAGAGTGGGAAGGAATCTGATGAGAAGTTT 4620
QY 438 ThrLeuPheLysGlu-----GlnLysLysAsn 446
Db 4621 AAACAGTATAAGAAATATTAGTCAGGTAGACATTAAAGAGTTGCAAGATGAAAGACTGAT 4680
QY 447 PheGluArgGlu-----ArgArgSerPheThrGluAlaAlaIleArgLeuGly 462
Db 4681 TTGGAATCTCAATCGTAGAATTCAGAAAAAATCGAAGAGCTACAAGTGAA----- 4734
QY 463 LeuGluArgLysAlaPheGluGluArgAlaSerTrpValLysGlnGlnPheLeuAsn 482
Db 4735 CTGAAAGACGAAAGCAGAAATCTAACAGTAGATAGACACCATACAGCAAGAGCTAACTAAT 4794
QY 483 MetThrAsnPheAspHisGlnAsn---SerGluAsnValLysLeuPheSerAlaPheSer 501
Db 4795 GCTTATGAGAAATAAGAGTTAATGAAGAGAAATGTATTATTA----- 4839
QY 502 GlySerSerAspProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeuHis 521
Db 4840 ---AATCCAAATTGAAATAATTAGAAATCAATTTAAGAGATAATCAAGACGAGATACAG 4896
QY 522 SerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAla--- 540
Db 4897 AGC-----ACTAAGGAAAGAAAGCAAGCAATCATTTGGATTCTTAC 4932
QY 541 -----SerProSerThrSerAspPheArgGlnThrHisSerCysValSerGlu 556
Db 4933 TTGAAAGATCTGAAGCAGAGTTGACTGACGCTCAACAAAGAGCTAAAGAAATCGCAAGAA 4992
QY 557 HisSerSerIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAla 576
Db 4993 AAGAGTAAACCGAAATTTGAACAACCTTCAAGCAGAGAGCGCTCTCGATCTCATGAATGCC 5052
QY 577 ArgGlu-----SerThrAspGlnLysTrp 584
Db 5053 AAACAACCTTGAATTAATAATACAGTAACCTTGAACCGAAGACAAACTTGG 5103
```

## RESULT 5

```
US-11-217-529-77055
; Sequence 77055, Application US/11217529
; Publication No. US20060099612A1
; GENERAL INFORMATION:
; APPLICANT: SUNTORY LIMITED
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHIISA
```

```
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: S-38-285
; CURRENT APPLICATION NUMBER: US/11/217,529
; CURRENT FILING DATE: 2005-09-02
; PRIOR APPLICATION NUMBER: US 10/932,182
; PRIOR FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 77055
; LENGTH: 5040
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-11-217-529-77055

Alignment Scores:
Pred. No.: 2,75e-11 Length: 5040
Score: 230.50 Matches: 147
Percent Similarity: 36.0% Conservative: 120
Best Local Similarity: 19.8% Mismatches: 230
Query Match: 7.3% Indels: 245
DB: 29 Gaps: 29

US-10-644-084-2 (1-615) x US-11-217-529-77055 (1-5040)

QY 15 GluAsnLysAsnLeuSerGlnTyrThr----- 23
Db 3028 GAAATTAAGATCTATCTGAATTTGGTTATCCGCTTGAAGAGTACGGCTGATGCCAA 3087
QY 24 SerGluThrLysMetSerProSerSerLeuTyrSerGlnGlnValLeuCysSerVal 43
Db 3088 GCAGAAATTAACAAAAACCAATCCCTCATTTGTATAGCACTCAGGATTTA----- 3135
QY 44 ProLeuSerLysAsnValHisGlyValPheGlyValPheCysThrGlyGlu----- 60
Db 3136 ---CTCGATTAACATGAAAGAGAAATGGATGGAGGAAAAAGGCTGATTTATGAAGAGAAATTTG 3192
QY 61 -----AsnIleGluGlnSer----- 68
Db 3193 ATATCTAATATTGAACAGACGGAATCCTTCGCTGTTGAAAAATTCGGTATTGATGAAAG 3252
QY 69 LeuAspGlnGluLeuThrThrPheGly-----Phe 78
Db 3253 GTTGATGACACAGCAGCAAAATAATGCGATAAAGATCATTTGAAGCTAGTCTTTCTTTGTTT 3312
QY 79 ProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAla 98
Db 3313 TCAAACTTACGTCATGAAAGAAATTTCTTTAGAGACGAAAG----- 3351
QY 99 ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuAlaGlnGlu 118
Db 3352 TTAACCACTGCAAGAGAGAAATTCGATTCGTTAAACAAAGACGATAGTTTAGAAAAA 3411
QY 119 SerVal---GluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCys 137
Db 3412 ACGATAAATGACCTCAACGAAACCCAGACTTTTATCGAGAGGAATAC-----CAATGC 3465
QY 138 TyrAlaLysLeuLysGluGlnLeuThrSerArgArgGluMetIleGly----- 154
Db 3466 TCTGCGATTATTATTGACGAATTTCAAGGATATACGAAAGAAAGTTACTCAGGTTAACATT 3525
QY 155 -----LeuGlnGluArgAspArgGlnLeuGlnCysLysAsnArg 167
Db 3526 TTAAGAGAGATAATGCCATCTTACAAAGAGTCTCAAAAGAGTAAACCGAAAAAACAGG 3585
QY 168 SerLeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleIle 187
Db 3586 GAAATTTATAACAATTAATGATAGGCAAGAGAAATTTCTCGATTACAGCGTGATCTC 3645
QY 188 AlaSerArgAlaThrGlnTyr-----AsnHisAspValLysArgLysGluArgGlu 204
Db -----AsnHisAspValLysArgLysGluArgGlu 204
```



```

Db 40 ---CAAGATGCTCGTACTAGATGATTAATTCAGAGTTGAAAAGGAAATCAATCTTCC 96
Qy 96 ILeValala-----ValLeuAsnCysMetAsnGluLeuLeuValLeuGlnArg 111
Db 97 ATTATTCAACAGCATGACGACATTATGGAGAAATTAATCAATTAATCTATTAAAGAGAA 156
Qy 112 LysAsnLeuLeuAlaGlnGlnSerValGlnThrGlnAsnLeuLys-----LeuGly 128
Db 157 AGTAACATACATTCGCGAATGAGCTGGAAAAACAACAATAACAGAGAAAGGAACATGCAA 216
Qy 129 SerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGluThrSer 148
Db 217 TCTGAATTAGATAAATTGAACCAAAATGTTGCCTTATCGAGTCCGAATTGATCGCTTG 276
Qy 149 ArgArgGluMetIleGlyLeuGlnGlnArgAspArgGlnLeuGlnCysLysAsnArgSer 168
Db 277 AAATATTCTATG-----CAAGAAAAAGCAAGAGAGCTCAATATTAGCTAAAGAGAG 327
Qy 169 LeuHisGlnLeuLeuLysAsnGluLysAspGluValGlnLysLeuGlnAsnIleAla 188
Db 328 GTTCATCGTTGGAAAAAGCGCTCACAGACATATTGGAGAAACATGAACAATTTGAGCTCA 387
Qy 189 SerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGluTyrAsnLysLeu 208
Db 388 AGC-----GATTATGAGAGCTAGAAAGCGAGATAGAAATTTG 426
Qy 209 LysGluArgLeuHisGlnLeuValMetAsnLysLysAspLysAsnIle---AlaMetAsp 227
Db 427 AAGGAGGAACCTAGAA-----AATAGGAGCGTCAAGGAGCGGAAGCGGAGAA 474
Qy 228 ValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLysThr 247
Db 475 AAATTTAAGCGGTGAGAACACGCGCAGGAGAGA-----TTAAAAACATCAAAATC 528
Qy 248 GluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArgGln 267
Db 529 TCACAGGACTCATGACTGACNAAGTAAGTAAATAGTCTAAGGGATGCAAGAACGTTGGAA 588
Qy 268 LysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnGlnMetLysLysGlu 287
Db 589 AATTCCTTGAGTGAGCAAAACGCGAGAAATCGAAGAG---TTACAAAATCAAAAGTA--- 642
Qy 288 MetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThr 307
Db 643 -----GCACAAGGTAAACAAC 657
Qy 308 GlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerVal 327
Db 658 CAGTTAGAGCAATAGAAAATTACAAAGAGCGCAAAAGCGTTCAAGAGAGCTTCAA 717
Qy 328 TrpGlyLeuSerCysAspThrValArg---GluGlnLeuThrAsnSerIleArgLysGln 346
Db 718 GCCAAGTTAGAGAAAGTACGACTCTTACGAATCTACGATAAAGCGCTTAAATGAAGAA 777
Qy 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366
Db 778 ATTACAACATTTAAAGAGAAATTTGAAAAACAAAGGCAAAATCCAGCAACAGTTACAAGCT 837
Qy 367 GluGlyLeuAsnGluGluAspValIleSerArgGlnAspHisGluGlnGlnThrGluLys 386
Db 838 ACATCTGCAATAGCAAAATGACTTATCT-----AACATAGTTGAATCTATGAAA 888
Qy 387 LeuGluLeuGluIleGluCysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGln 406
Db 889 AAGTCTTTTGAAGAGATATAAATCAAA---TTTCATCAAGAAAAAACCACCAAGAGTTAAT 945
Qy 407 GlnGlnLeuAlaThrThrCysAspAspAspThrThrSerLeuLeuArgAspCysTyrLeu 426
Db 946 GAAAAA-----ATA 954
Qy 427 LeuGluGluLysGluArgLeuLysGluGluTyrThrLeu---PhelLysGluGlnLysLys 445

```

```

Db 955 CTTGAGGCCCAAGAAAGGCTTAATCAACCTTCCAATATCAATATGAGGAGATTAATAAAA 1014
Qy 446 AsnPheGluArgGluArg----- 451
Db 1015 AAATGGGAATCTGAGCAACAGCAAGAGTATCTCAAAAAGATTCCGCAAGCTGAGGAAGCC 1074
Qy 452 -----ArgSerPheThrGluAlaIleAArgLeuGlyLeuGluArgLys 466
Db 1075 CTCAAAAAGCGAATCAGATTACCCACTGAGGAGAAAATCAATAGATAATTCGAAACGAAAG 1134
Qy 467 AlaPheGluGluGluArgAlaSerTrpValLysGlnPheLeuAsnMetThrAsnPhe 486
Db 1135 AAGGAGCAATTTGGAAAA-----GAGTTT 1158
Qy 487 AspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSerAsp--- 505
Db 1159 GAAGAAAAGGTTGAGGAGAGATAAATCAATGGAACAA---TCTGGAGAAATAGACGTG 1215
Qy 506 -----ProAspAsnLeuIleValHisSerArgProArgGlnLysLysLeuHisSerVal 523
Db 1216 GTGCTTCGAAAAACAGCTAGAGCTAAGGTTCAAGAGAGAAACAAAGGAATTTGGAACGAG 1275
Qy 524 AlaAsnGlyValProAlaCysThrSerLysLeu-----ThrLysSerLeuProAla 540
Db 1276 TATAC-----AAAAATTACAAGAAAGAACTCAAGATGTACCACAC 1317
Qy 541 SerProSerThrSer----- 545
Db 1318 TCCCCTCATATCTCAGGTGATGAGAGGAGCAAGTTACGGGCGAGAAATCGAAAGCAGTTG 1377
Qy 546 -----AppPheArgGlnThrHisSerCysValSerGluHisSer----- 558
Db 1378 AGGAGGAGGTTCAACAATGAACCTGCAAGCCATAAAGAGAAATCTCTTCGACGAGGAAG 1437
Qy 559 SerIleSerValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGlu 578
Db 1438 CAACAGCAATGATGAAAACCTACCTTTTGGAAAGAAACTTGCCTCAAGATGGATCTCAA 1497
Qy 579 SerThrAspGlnLysTrpSerValGlnSerArgProSerSer 592
Db 1498 TTGTGAGAAACAAACAAAGTCCGAGACTCTCCGAAATCT 1539

RESULT 7
US-11-293-697-1966
; Sequence 1966, Application US/11293697
; Publication No. US20060105376A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/11/293,697
; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1966
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-293-697-1966

Alignment Scores:
Pred. No.: 9,06e-10 Length: 2479
Score: 209.00 Matches: 108
Percent Similarity: 42.4% Conservative: 113
Best Local Similarity: 20.7% Mismatches: 192
Query Match: 6.6% Indels: 109
DB: 7 Gaps: 20

US-10-644-084-2 (1-615) x US-11-293-697-1966 (1-2479)
Qy 12 LeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSer 31

```





Db 2520 AGCA-----TTGGAAGCCAGAGCCAGTCCAGA 2546  
QY 543 rThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSerValle 563  
Db 2547 GATGAGTACCGAATACAG-----CACTTGGAGAGAGATCAC 2585  
QY 563 uAenIleThrProGluGluSerLysProSer-----GluValAl 576  
Db 2586 CAGGTACAAAGATGAATCTAGCAAGGCCAGGCGAAGTTGATCGACTCTTAGMAATCTT 2645  
QY 576 aArgGluSerThrAspGlnLys 583  
Db 2646 GAAGGAGGTGGAAAATGAGAAG 2667

RESULT 9

US-10-505-928-326  
; Sequence 326, Application US/10505928  
; Publication No. US20060088532A1  
; GENERAL INFORMATION:  
; APPLICANT: Ludwig Institute for Cancer Research et al.  
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES  
; FILE REFERENCE: 28967/39178  
; CURRENT APPLICATION NUMBER: US/10/505,928  
; CURRENT FILING DATE: 2004-08-27  
; PRIOR APPLICATION NUMBER: US 60/363,019  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: PatentIn 3.2  
; SEQ ID NO 326  
; LENGTH: 10211  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-505-928-326

Alignment Scores:  
Pred. No.: 1,09e-08 Length: 10211  
Score: 206.50 Matches: 156  
Percent Similarity: 35.8% Conservative: 107  
Best Local Similarity: 21.2% Mismatches: 241  
Query Match: 6.5% Indels: 231  
DB: 6 Gaps: 33

US-10-644-084-2 (1-615) x US-10-505-928-326 (1-10211)

QY 17 LysAsnLeuSerGlnTyrThr----- 23  
Db 1657 AGAATACTCTCAGAAACCATGTTAAGAGATCTTCAAGAAAAATAAATCAGCAAGNA 1716  
QY 24 -----SerGluThrLysMetSerProSerSerLysGlnVal 38  
Db 1717 AACTCTTGACTTTAGAAAACTGAAGCTTGCTGGCTGATCTGGAAAAAGCAGCGAGAT 1776  
QY 39 LeuCysSerSerValProLeuSerLysAsnValHisGlyValPheGlyValPheCysThr 58  
Db 1777 ---TGTTCTCAAGACCTTTGAAGAAAGAGAACATCACTT----- 1815  
QY 59 GlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPhe 78  
Db 1816 ----GAACAATTAATGATAGTTAAGCAAGACAGAGAAAGATGCCAAAGCCTTG---CTG 1869  
QY 79 ProSerLeuTyrGluGluSerLysGluAlaLysArgGluLeuAsnIleValAla 98  
Db 1870 AGTGCTTTAGAGTTAAAAAGAAAGAAATATGAATTGAAAGAGAGAAAACTCTGTTTCT 1929  
QY 99 ValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGlu 118  
Db 1930 TGTGGAAAAAGTGAAACCGAA-----AAACTTTTAACACTCAGATG 1968  
QY 119 SerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyr 138  
Db 1969 GAATCAGAAAGAAACCTTGAG-----AGTAAATTAATCACTTGGAAACTTGT--- 2019  
QY 139 AlaLysLeuLysGluGlnLeuGluThrSerArg-----ArgGluMetIleGlyLeuGln 156

Db 2020 -----CTGAAGACACAGCAAAATAAAAAAGTCATGAATACACAGAGAGTAAAGACGCTG 2073  
QY 157 GluArgAspArgGln---LeuGlnCysLysAsnArgSerLeuHisGlnLeuLeu----- 173  
Db 2074 GAGATGGACAGAAAAACCTAAGTGTGAGATCAGAAACCTTCACAACTGTGTACACAGT 2133  
QY 174 LysAsnGluLysAspGluValGlnLysLeuGlnHisIleAlaSerArgAlaThrGln 193  
Db 2134 AAGTCAGTGGAGGTAGAGACCAGAAAACCTAGCTTATATGGAGCTACAGCAGAAAGCTGAG 2193  
QY 194 TyrAsnHisAspValLysArgLysGlu----- 202  
Db 2194 TTCTCAGATCAGAAAAATCAGAGAAGAAATAGAAAAATATGTGTTTGAAGACTTCTCAGCTT 2253  
QY 202 ----- 202  
Db 2254 ACTGGGCAAGTTGAAGATCTAGAACACAAAGCTTCAGTTACTGTCAAAATGAATAATGGAC 2313  
QY 203 -----ArgGluTyrAsnLysLeuLysGluArgLeu--- 212  
Db 2314 AAAGACCGGTGTACCAAGACTTGCATGCCGAATATGAGAGCTTCAGGGATCTGTAAAA 2373  
QY 213 -----HisGlnLeuValMetAsnLysLysAsp 221  
Db 2374 TCCAAGATGCTTCTCTGGTGACAAATGAAGATCATCAGAGAAGTCTTTTGGCTTTTGTAT 2433  
QY 222 LysAsnIleAlaMet-----AspValLeuAsnTyrValGlyArgAlaAspGlyLysArg 239  
Db 2434 CAGCAGCCTGCATGCATCATCTCTTTGCCAATATAATT-----GGAGAACA 2481  
QY 240 GlySerTyrArgThrAspLysThrGluAlaArgAsnGluAsp----- 253  
Db 2482 GGAAGCATGCCTTCAGAGAGGAGTGATGTCGTTTAGAAGCAGACCAAGATCCGAAAAAT 2541  
QY 254 -----GluMetTyrLysIle 258  
Db 2542 TCTGCCATCTACAAAAATAGAGTTGATTCACTTGAATTTTCATTAGAGTCTCAAAAAACAG 2601  
QY 259 LeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGluLeuLys 278  
Db 2602 ATGAACCTCAGACCTGCAAAACGAGTGTGAGAGTGTGGCAATCAAAAGGAGAAATAGAA 2661  
QY 279 LysValLeu-----GlnGlnMetLysLysGluMetIleSerLeuLeuSerProGln 295  
Db 2662 GAAAACTCTCATGAAAGCAGACAGATGCATCAAAAGTTTGTGGCTGAAACAAAGTCAGCGC 2721  
QY 296 LysLysLysProArgGluArgAlaGluAspGlyThrGlyThrValAla-----IleSer 313  
Db 2722 ATTAGTAAGTTACAGGAGACACTTCTGCTCACCAGAAATGTTGTTGCTGAAACCTTAAGT 2781  
QY 314 AspIleGluAspAspSerGlyLysLeuSer-----ArgAspSerValTyr----- 328  
Db 2782 GCCCTTGAGAAACAAGGAAAAAGAGCTGCAACTTTTAATGATAGGTAGAACTCAGCAG 2841  
QY 329 -----GlyLeuSerCysAspThrValArgGlu--- 337  
Db 2842 GCAGAGATTCAGAATTAATAAAGAGCAACCATCTACTTGAAGACTCTCTAAAGGAGCTA 2901  
QY 338 GlnLeu-----ThrAsnSerIleArgLysGln-----TrpArgIle 349  
Db 2902 CAACCTTTATCCGAAACCCCTAAGCTTGGAGAGAAAGAAATGAGTTCCATCATTTCTCTA 2961  
QY 350 LeuLysSerHisValGluLysLeu-----AspAsnGlnAlaSerLysValHisSerGlu 367  
Db 2962 AATAAAGGGAATTTGAAGAGCTGACCCAGAGAAATGGGACTCTTTAAGGAAATTAATGCA 3021  
QY 368 GlyLeuAsnGluGluAspVal----- 374  
Db 3022 TCCTTAATCAAGAGAGAGATCACTTAATCCAGAAAAAGTGAGAGTTTGTCAAACTATATA 3081  
QY 375 -----IleSerArgGlnAspHisGluGlnGluThrGluLysLeuGlu 388



Db 3082 GATGAAAGGAGAGAAAGCATTTTCAGAGTTATCTGATCAGTACAGCAAGCAAGAAACCTTATT 3141  
 Qy 389 LeuGluIleGluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGlnGln--- 407  
 Db 3142 TTA---CTAAAGATGTGAAGAAACCGGAAATGCGATATGAGGATCTTTAGTCAAAATAC 3198  
 Qy 408 GlnLeuAlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeu 427  
 Db 3199 AAAGCAGCAGCAAGAAAGAAATCTTAATTAAGATGCTTGCTAAATGAATGCATGACTGCTT 3258  
 Qy 428 GluGluLysGluArgLeuLysGluGluTyrThrLeuPheLysGluGlnLysLysAsnPhe 447  
 Db 3259 TGTGAAATAGGAAATAAGTGGTGA---CAGCTAAAGGAAGCATTT 3303  
 Qy 448 GluArgGluArgSerPheThrGluAlaAlaIleArgLeuGluArgLysAla 467  
 Db 3304 GCAAAGGACACCAAGAAATTC-----TTAACAATAATAGCA 3339  
 Qy 468 PheGluGluGluArgAlaSerTrpValLys-----GlnGlnPheLeu 481  
 Db 3340 TTTCCTGAAGAAAGAAATCAGAAATCTGATCTAGATTGGAGACAGTGCAGCAAGCTCTG 3399  
 Qy 482 AsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSer 501  
 Db 3400 AGATCTGAGATGACAGATAACCAAAACAATCTTAAG-----AGC 3438  
 Qy 502 GlySerSerAspProAsnLeuIleValHisSerArgProGlnLysLysLeuHis 521  
 Db 3439 GAGGCTGGTGGTTAAAGCAAGAAATCATGACTTTAAAGGAAGCAAAACAATAATGCAA 3498  
 Qy 522 SerValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSer 541  
 Db 3499 AGGAAGTTAATGACTTTATACAGAGATGAACAGCTGATGAAGTAATG----- 3549  
 Qy 542 ProSerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIleSer 561  
 Db 3550 -----AAGACTAAACATGAATGT----- 3567  
 Qy 562 ValLeuAsnIleThrProGluGluSerLysProSerGluValAlaArgGluSerThrAsp 581  
 Db 3568 -----CAAAATCTAGAATCAGAACCAATTAGGAAC----- 3597  
 Qy 582 GlnLysTrpSerValGlnSerArgProSerSerArgGluGluCys 596  
 Db 3598 -----TCTGTGAAGAAAGAGAGAGTGAAGAAATCAATGT 3633

## RESULT 10

; US-11-217-529-6026  
 ; Sequence 6026, Application US/11217529  
 ; Publication No. US20060099612A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SUNTORY LIMITED  
 ; APPLICANT: NAKAO, YOSHIHIRO  
 ; APPLICANT: NAKAMURA, NORIHIRO  
 ; APPLICANT: KODAMA, YUKIO  
 ; APPLICANT: FUJIMURA, TOMOKO  
 ; APPLICANT: ASHIKARI, TOSHIHIKO  
 ; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
 ; FILE REFERENCE: S-38-285  
 ; CURRENT APPLICATION NUMBER: US/11/217,529  
 ; CURRENT FILING DATE: 2005-09-02  
 ; PRIOR APPLICATION NUMBER: US 10/932,182  
 ; PRIOR FILING DATE: 2004-09-02  
 ; NUMBER OF SEQ ID NOS: 197023  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 6026  
 ; LENGTH: 5658  
 ; TYPE: DNA  
 ; ORGANISM: Saccharomyces pastorianus  
 ; US-11-217-529-6026

Alignment Scores:

Pred. No.: 6.58e-09 Length: 5658

Score: 205.00 Matches: 150  
 Percent Similarity: 37.6% Conservative: 118  
 Best Local Similarity: 21.1% Mismatches: 236  
 Query Match: 6.5% Indels: 208  
 DB: 7 Gaps: 35  
 US-10-644-084-2 (1-615) x US-11-217-529-6026 (1-5658)  
 Qy 12 LeuCysThrGluAsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSer 31  
 Db 1366 TTAACGCTCAAAATGAAACAACTGGCGGAATATGAAGATGAATCCACATA----- 1416  
 Qy 32 SerLeuTyrSerGlnGlnVal---LeuCysSerSerVal----- 43  
 Db 1417 ---TTGTCCAAACAACGCTCTGATTTATGTCGTCAAATATCAATATCTCTTGTATCACAAT 1473  
 Qy 44 -----ProLeuSerLysAsnValHisGlyValPheGlyVal 55  
 Db 1474 TCAGTATCCAATGACTCTTAAGGACCTTTACCCAA----- 1509  
 Qy 56 PheCysThrGlyGluAsnIleGlu-----GlnSerIleSerTyrLeuAspGlnGluLeu 73  
 Db 1510 -----GAGAAATCAAAATTTATTCAAAACATTTTACAAACCGCATAGTCTGCT 1557  
 Qy 74 ThrThrPheGlyPheProSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGlu 93  
 Db 1558 ACAACT-----GAATCAGACTCGCAAAAAAATTTGTAACGGAA 1593  
 Qy 94 LeuAsnIleValAlaValLeuAsnCysMetAsnGluLeuValLeuGlnArgLysAsn 113  
 Db 1594 ---AGTTAGTTGAATTTAGAAAT-----ATTATTCAGTTACAGAGAAAAAT 1638  
 Qy 114 LeuLeuAlaGlnGluSerValGluThrGlnAsnLeuLysLeuGlySerAspMetAspHis 133  
 Db 1639 ---TCAGATTCTTGAGAGTAAACAAGAACTTA-----GCCGATAAA 1677  
 Qy 134 LeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuGlu----- 146  
 Db 1678 TTAGAGTCAAAATGAAATAAATCTAAACAATATCTTAAGAAGATTGAAACCAAACTATA 1737  
 Qy 147 ---ThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLys 165  
 Db 1738 AATGAGCTAAAGACGATCCTTACCTTTACAAAGCCAAAAAGCACAGTTAGTAATCGAA 1797  
 Qy 166 AsnArgSerLeuHisGlnLeuLeuLysAsnGluLys-----Asp 178  
 Db 1798 -----ATCCACGAGTTAGAAAGGAACGAGAAAAATTTAAGGACTGGACTTTGGAT 1848  
 Qy 179 GluValGlnLysLeuGlnAsnIleAlaSerArgAlaThrGlnTyrAsnHisAspVal 198  
 Db 1849 CAAAAGGCTTCACCTAATACTCGATAATAACAATACTAAGACTGAGACTAAAAAGAGAGCTA 1908  
 Qy 199 LysArgLysGluArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsn 218  
 Db 1909 -----GAATCCCAATCCAAGACTTACAGCCCGCTATTCTCAAGTTACCGAGAA 1959  
 Qy 219 LysLysAspLysAsnIleAlaMetAspValLeuAsn-----TyrVal 232  
 Db 1960 TCTACCGAGAAT-----ATGTCACCTTTGAATAAGGAATTACAGGACCTATATGAT 2010  
 Qy 233 GlyArgAlaAsp-----GlyLysArgGlySerTrpArg----- 243  
 Db 2011 GGCAAAAGTAGCCTATCAATTGAGCTAGGAGGAAAGTCACTCTAGGATATTGGCAGAA 2070  
 Qy 244 -----ThrAspLysThrGluAlaArgAsnGluAspGlu 254  
 Db 2071 GAGAGTTCAAAATTAATTTCAATACGCTAGATTGACTTAAGCCGAGAAC---GACCAA 2127  
 Qy 255 MetTyrLysIleLeuLeuAsn-----AspTyrGluTyrArgGlnLys 268  
 Db 2128 TTACGCAAGAGAGCAGTTAATTTACAAAATGCCATCTCAAAAACAGATTCAAGAGCCAG 2187  
 Qy 269 GlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLysGluMet 288

Db 2188 GAGACACTTAATGAGTACGTTTCTTGCATAATCGAAACTAAGTGTACTGAAACAGAAATTA 2247  
Qy 289 ILeSerLeuLeuSerProGln-----LysLysLysProArgGlu 301  
Db 2248 TCCAATTTGAAGTCAGACGACACTGAAATTCGAATTTAGAAAAAAACCCTAAGACAGAAA 2307  
Qy 302 ArgAlaGluaspGlyThrGlyThrValAlaIleSerAspIleGluAspSerGlyGlu 321  
Db 2308 -----CTAAGTGAACCTTCTTCTGAGAAAAACAGT 2337  
Qy 322 LeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAen 341  
Db 2338 TTACATATATGTTAACTCAATTCGAAACCTCCAAAAGGAACGTCGAGAACCTATTAGAT 2397  
Qy 342 SerIleArgGlnTrpArgIleLeuLysSerHisValGluLysLeuAspAenGlnAla 361  
Db 2398 GAATTTAAATCT-----TGTCAAAACAAATAGGTATATTAGAAATCGGMA 2448  
Qy 362 SerLysValHisSerGluGlyLeuAenGluGluAspValIleSer----- 376  
Db 2449 AATGAACATAAAACCGAGCTATTTCATAAAGCAACATACATTAAACAACTGGAAGAAGAT 2508  
Qy 377 -----ArgGlnAspHisGlu 381  
Db 2509 AATGACACAAAAGTTGAATGGTATCAGAACAAAATTTGAAACTTTCGAGGAAGATCATGAA 2568  
Qy 382 Gln-----GluThrGluLysLeuGluLeuGluIle 391  
Db 2569 TCTATTATGAATCTTTAAATGAAGAAAGCAAAATTTGAGTTGAGAAATTCGAGTATGAAGTC 2628  
Qy 392 GluArgCysLysGluMetIleLysAlaGlnGlnLeuLeuGln-----GlnGlnLeu 409  
Db 2629 AAGCTTTAGAAAAGGAATCGAAGAAACAGATTCGTTTACATCTTACATCTACTA 2688  
Qy 410 AlaThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrIleuLeuGluGlu 429  
Db 2689 GATGAATCTGTTAAACGATGAT-----TCCTTACGCAGAGAA-----CTGGAAGAAGTCG 2736  
Qy 430 LysGluArgLeuLysGluGlnTrpThrLeuPheLysGluGlnLysLysAenPheGluArg 449  
Db 2737 AAAATAAACTTAACAGACTTATCTCAATCCAGATAACAAAAGCTTTTACGAGACT 2796  
Qy 450 GluArgArgSerPheThrGluAlaAla----- 458  
Db 2797 GCTGATAATCTCTACAGGAATGACATCCATCAAGATGAATCTAATAAATCATTTTCT 2856  
Qy 459 -----IleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpVal 476  
Db 2857 AATCAAAATTCAAAGTTTAACTGATGAAAAGACCACTTTTAGAGGATATAAGTTTCTCTG 2916  
Qy 477 LysGlnGlnPheLeuAenMetThrAen---PheAspHisGlnAenSerGluAenValLys 495  
Db 2917 AAAGAGCAATGCTTAATTTGAATATGATGATTTTACAATAACAGCCATGGAATA 2976  
Qy 496 LeuPheSerAlaPheSerGlySerSerAspProAspAenLeuIleValHisSerArgPro 515  
Db 2977 GAAAAATCCGAATTCAG-----AAAAAGATTTCATTATTTTACAAAT 3018  
Qy 516 ArgGlnLysLysLeuHisSerValAlaAenGlyValProAlaCysThrSerLysLeuThr 535  
Db 3019 AACACAAAGAAATTTGAAGCGGTCAAAATCTGTAGTATGAGTCCAAAGTTGTGGAAGATTCAA 3078  
Qy 536 LysSerLeuPro-----AlaSerProSerThrSerAspPheArg----- 548  
Db 3079 AAGGATCTTGACAGCAAACTACATACGCTAATACTATGCCCCAAAATAACTATGAACAGGAA 3138  
Qy 549 ----GlnThrHisSerCysValSerGlu----- 556  
Db 3139 TTGCAAAAACACGACAGATGTTTCCAAGACTATTAGTGAACATAAGAGAAACAATTACATACT 3198  
Qy 557 ----HisSerSerIleSerValLeuAenIleThrProGluLeuSerLysProSerGluVal 575

Db 3199 TACAAGAGCTCAAGTAGAAACCTTTAAACCTAACACGTGACCAG-----TTTGAAAAA 3249  
Qy 576 AlaArgGluSerThrAspGlnLysTrpSerValGln 587  
Db 3250 ACTCTAAGGAAAGCAAGAAACAATTTGGAATTTCCCA 3285  
RESULT 11  
US-11-529-76923  
; Sequence 76923, Application US/11217529  
; Publication No. US20060099612A1  
; GENERAL INFORMATION:  
; APPLICANT: SUNTORY LIMITED  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUJIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: S-38-285  
; CURRENT APPLICATION NUMBER: US/11/217,529  
; CURRENT FILING DATE: 2005-09-02  
; PRIOR APPLICATION NUMBER: US 10/932,182  
; PRIOR FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 76923  
; LENGTH: 5787  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-11-529-76923  
Alignment Scores:  
Pred. No.: 1,41e-08 Length: 5787  
Score: 201.50 Matches: 132  
Percent Similarity: 37.9% Conservative: 97  
Best Local Similarity: 21.9% Mismatches: 215  
Query Match: 6.4% Indels: 160  
DB: 25 Gaps: 25  
US-10-644-084-2 (1-615) x US-11-217-529-76923 (1-5787)  
Qy 60 GluAenIleGluGlnSerIleSerTyrIleuAspGlnLeuThrThrPheGlyPhePro 79  
Db 2590 AAAAAATTCACAGCAGCAAAATTAATAAATGAGAACCGACCTTCAAGAAATGGAATCTTAAG 2649  
Qy 80 SerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeu---AsnIleValAla 98  
Db 2650 AAGAAGTTTTTGGAGAGAAAGAACCAAAACTGTCAATGAGTTGGAAAAATACGCAGGAC 2709  
Qy 99 ValLeuAenCysMetAenGluLeuValLeuGlnArgLysAen-----LeuLeu--- 115  
Db 2710 TTGCTAAATCAGGAGAAAGAGCAATTTA-----AGAAAAAATGAGAGCTTATTAAT 2760  
Qy 116 -----AlaGlnGluSerValGluThrGlnAenLeuLysLeuGlySerAspMet 131  
Db 2761 AGAGTTAAACACCTCAGAAACATTAACAAAACATTCGATGACCTTTGTTCTGAAAG 2820  
Qy 132 AspHisLeuGlnSerCysTyrAlaLysLeuLysGluGlnLeuThrSerArgArgGlu 151  
Db 2821 GATGAAATTCACAGGAAAAAACTAGAGTTCCGCAAAATCTTTGAAGAGCTCATCAAAA 2880  
Qy 152 MetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCysLysAenArgSerLeuHis--- 170  
Db 2881 ATCCAAGCCCTTCAAGAAACTATTAGAAACGGGCGGACCTTAGAGAAATTTACACTCA 2940  
Qy 171 -----GlnLeuLeuLysAenGluLysAsp-----GluValGln 181  
Db 2941 AAAAAATACGAACATAATAAAGCAAAATATCAGACTTGAACCTGTGATATTCTTAAGGAACAG 3000  
Qy 182 LysLeuGlnAenIleIleAlaSerArgAlaThrGlnTyrAenHisAspValLysArg--- 200  
Db 3001 AGCTCCCAAGTCATTGATCAAGAAATCTAAATGAGCTGGAAAAAGAAATAAGAGATT 3060

```
QY 201 -----LysGluArgGluTyrAsnLysLeuLysGluArgLeuHis--- 213
Db 3061 AAGATGTCATAAATCGAAGAGGAGAAATAAATCGTTCAACGATAGTTAAAGTTCC 3120
QY 214 -----GlnLeuValMetAsnLysLysAspLysAsnIleAlaMet 226
Db 3121 TCAGAGAAGATCTTGACATAAAATTTGGTCACITTTAGAGAAAAATTTGTAACATTTGCGATG 3180
QY 227 AspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGlySerTrpArgThrAspLys 246
Db 3181 TCAAGACTACAA-----TCCCTTGTAAACAGAAAAAT 3210
QY 247 ThrGluAlaArgAsnGluAspGluMetTyrLysIleLeuLeuAsnAspTyrGluTyrArg 266
Db 3211 TCAGATTGGTTTCGAAATAATGAGAAATTTCAAG----- 3243
QY 267 GlnLysGlnIleLeuMetGluAsnAlaGluLeuLysLysValLeuGlnMetLysLys 286
Db 3244 -----AAAGAAAAGGCAGCACTGAAT-----AACCAGTTGAAAAAT 3279
QY 287 GluMetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGly 306
Db 3280 AATGAAAGTCAACTGTTAAGATGAAGAAAAAATTTGATAATCACAAGAAAAAGAGCTTGCT 3339
QY 307 ThrGlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSer 326
Db 3340 ACTTTTCCAAGCAAGAGATCAGCGAGTAGTGAACATGGTAATAATTACTGCTGAA--- 3396
QY 327 ValTrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGln 346
Db 3397 -----CTAAGAGACAGAATCCAACCTTACCAG----- 3426
QY 347 TrpArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSer 366
Db 3427 -----TATAAATCTAATCAAAAGATTAAAGAGAATATATTTCCAAC----- 3468
QY 367 GluGlyLeuAsnGluAspValIleSerArgGlnAspHisGluGlnGluThrGluLys 386
Db 3469 -----TTCCAAGAGAAAAACGAGGAAACAGAGCAAAAAGAAA 3504
QY 387 -----LeuGluLeuGluIleGluArg 393
Db 3505 AGAAACAGCCTGGTGGTCTCTGAACGATAGTAAATAATTAAAGAAATTTGAA---GCTCGG 3561
QY 394 CysLysGluMetIleLysAlaGlnGlnGlnLeuLeuGlnGlnLeuAlaThrCys 413
Db 3562 TTGTACACAGGAATAATCTTAAATCAATCTTAAACAAAAGAAATATCAGGAAATTTCTGTA 3621
QY 414 AspAspAspThrThrSerLeuLeuArgAspCysTyrLeu-----LeuGluGlu 429
Db 3622 GAAACGAATATTTCGTCTACTAGACGTTCCACATCGTATTCAGATGATCCACTTGACAAA 3681
QY 430 LysGluArgLeuLysGluGluTrpThrLeu-----PheLysGluGlnLysLysAsn 446
Db 3682 GAAGATATAATAAAAAAATACTATGACCTTCAATTTAGCGTTTACTGAAATAACTAGGAAT 3741
QY 447 PheGluArgGlu-----ArgArgSerPheThrGluAla 457
Db 3742 CTAGAGAATGAAATTGAAGAGAAGAAGAACTTAATTTCTAGATTGAGATTTCACTGAAACA 3801
QY 458 AlaIleArgLeuGlyLeuGluArgLysAlaPheGluGluGluArgAlaSerTrpValLys 477
Db 3802 AGA-----CTAGCATCTTCGTCTTTTGAGGACCAAAAG-----ATTAAG 3840
QY 478 GlnGlnPheLeuAsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPhe 497
Db 3841 GCACAAATGAAGAAATTAATAAATTTGATCCAGGATATCGACCCCTAGTATTCTTTGGAC 3900
QY 498 SerAlaPheSer-----GlySerSerAspProAsnLeu 509
Db 3901 AGTATTTTAAATCAGCCGCTAAATAACTGCTCCCTGACAAAGAGTCTGATATTACAAATTA 3960
QY 510 IleValHisSerArgProArgGlnLysLysLeu-----HisSer 522
```

```
Db 3961 ATGCTTGAGCTCGATTATTATAAAGACAATTTGGATATTCGAACAAGAGCTCACTACGAT 4020
QY 523 ValAlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerPro 542
Db 4021 GCAGAAAAATGCCATATCTCTTTACACAGATAAATTTAGAAAG----- 4062
QY 543 SerThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerSerIleSer-Va 562
Db 4063 -----ATCCAAGAGGAAAGCTCCCTGTCTCATCT 4089
QY 562 IleAsnIleThrProGluGluSerLysProSerGluValAlaAlaArgGluSerThrAspG 582
Db 4090 TCTGATATTTACAAACTGAAGTTCGAAGCCA-----GTGAAGAAAGAGTCAAAATCCTTG 4143
QY 582 nLysTrpSer 585
Db 4144 GAAGACAAGC 4153

RESULT 12
US-11-289-102-97
; Sequence 97, Application US/11289102
; Publication No. US2006012151A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Hyerim
; APPLICANT: Shaw, Peter M.
; APPLICANT: Clark, Edwin
; TITLE OF INVENTION: BIOMARKERS AND METHODS FOR DETERMINING SENSITIVITY TO
; TITLE OF INVENTION: MICROTUBULE-STABILIZING AGENTS
; FILE REFERENCE: 10338 NP
; CURRENT APPLICATION NUMBER: US/11/289,102
; PRIOR FILING DATE: 2005-11-29
; PRIOR APPLICATION NUMBER: US 60/631,993
; NUMBER OF SEQ ID NOS: 395
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 97
; LENGTH: 4925
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-289-102-97

Alignment Scores:
Pred. No.: 8,15e-08 Length: 4925
Score: 192.00 Matches: 111
Percent Similarity: 40.5% Conservative: 112
Best Local Similarity: 20.2% Mismatches: 221
Query Match: 6.1% Indels: 107
DB: 7 Gaps: 18

US-10-644-084-2 (1-615) x US-11-289-102-97 (1-4925)
QY 11 ValLeuCysThrGluAsn-----LysAsnLeuSerGlnTyrThr 23
Db 1507 CTGGTATGCTTAAACAACACTGAGATTTCAGAGAACAGCTCTGACCTAGCCAGAAACTT 1566
QY 24 SerGluThrLysMetSerProSerSerLeuTyrSerGlnValLeuCysSerVal 43
Db 1567 AAAGAACTCAGAGCAAAATACAGGAGGC-TATGAAGAAGCTCTTAGTGTGCAGAGCA 1625
QY 44 ProLeuSerLysAsnVal-----HisGlyValPheGlyValPheCysThrGly 59
Db 1626 GATGAACCTCGCTCTGTCTCACCTGAAGCATGTGATAATTATTACATTTTCCAGAGCT 1685
QY 60 GluAsn-IleGluGlnSerIleSerTyrLeuAspGlnGluLeuThrThrPheGlyPhePr 79
Db 1686 GAGGTCACGGAAGAGGAAATAAATGTCTAAAGCAGGATCTGCAGAAAT----- 1734
QY 79 oSerLeuTyrGluGluSerLysSerLysGluAlaLysArgGluLeuAsnIleValAlaVa 99
Db 1735 -GCATTAGAAGAAAGTGAAGAAATAAAGAGAAAGTGAAGAGTTA----- 1779
QY 99 IleuAsnCysMetAsnGluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSe 119
```



```
QY 111 ArgLysAsnLeuLeuAlaGlnGluSerValcluThrGlnAsnLeuLysLeuGlySer--- 129
Db 3214 ACAGATCATCTAGTGTGAAATTCAGTGTAGTCGAAATATATGGATGATGGTGGC 3273
QY 130 -----AspMetAspHisLeuGln-----SerCysTyrAlaLysLeuLysGlu 143
Db 3274 GAAAGTAACGGTGATAAAGATATTTAAAGCTAGTTTCATTTATTTCCAGCTCGCGTCAT 3333
QY 144 GlnLeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGln 163
Db 3334 GAAAGGAATACTCTGGAACAACTA-----ACTACC 3366
QY 164 CysLysAsnArgSerLeuHisGlnLeuLeuLysAsnGlnLysAspGluValGlnLysLeu 183
Db 3367 TGTAG---AGAGATTTA---GCATTCGAAAGCAAAAGAAATGCCAATTTGGAAAGACC 3420
QY 184 GlnAsnIleIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArgLysGlu--- 202
Db 3421 GTAATGATATG-----CAGAAACTCACCCCATATCGCGCAGGATGTC 3465
QY 203 -----ArgGluTyrAsnLysLeuLysGluArgLeuHisGlnLeu 215
Db 3466 CAATGTTCTACCGAAATATTTGACGAGTTTGAGGATATTTATGAAGAAATAACTCAAGTT 3525
QY 216 ValMetAsnLysLysAspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAla 235
Db 3526 AACATCTTAAGGAACAAACACCATTTTCGGAAGAGCCTTGAGAGAGTT----- 3576
QY 236 AspGlyLysArgGlySerTrpArgTrpAspLysThrGluAlaArgAsnGluAspGluMet 255
Db 3577 -----ACAGAAAGAAATGAGCA-----ATT 3597
QY 256 TyrLysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAla 275
Db 3598 TATAAA-----GAGTACACTAATTTGCAACATGAATATCTCGACTACAAGGT 3645
QY 276 GluLeu-----LysLysValLeu----- 281
Db 3646 GAACCTGTGCAACGAAAGCAAACTTCGTAAATGCTTAATAAGGTTTTAGTCTATGAA 3705
QY 282 -----GlnGlnMetLysLysGluMetIleSerLeuLeuSerProGlnLysLysLys 298
Db 3706 AGTGAATTTGAACAATGGAACAGAGGTTTCAGAGCTTATCTCAACAACAGAAAGAGGCA 3765
QY 299 ProArgGluArgAlaGluAspGlyThrGlyThrValAlaIleSerAspIleGluAspAsp 318
Db 3766 CACAAAGAAGAACTGAAAAG-----CTATTTAATGAGATTAGCGACATG 3810
QY 319 SerGlyGluLeuSerArgAspSerValTrpGlyLeuSerCysAspThrValArgGluGln 338
Db 3811 AGGGCTAAACTT-----CTCAATGCCCCAGAAATGCAATGCTGAC 3849
QY 339 LeuThrAsnSerIleArgLysGlnTrpArgIleLeuLysSerHisValGluLysLeuAsp 358
Db 3850 TTGAATGATAAATTCAT-----AGGTTAAAGAAACACAGCGCAGCAAAAGTTGGAC 3900
QY 359 AsnGlnAlaSerLysValHisSerGluGlyLeu---AsnGluGluAspValIleSerArg 377
Db 3901 -----GCTTCCAGAAACAAACAACTGCTTGAACCACTTCAATGAATTTAAAGAT 3954
QY 378 GlnAspHisGlu-----GlnGluThrGlnLysLeuGluLeuGlu 390
Db 3955 ATGAGAAATGAATTTGGAAGAGAGTTTGGTAGTGAAGAATCTAAAGTCTCTGAACGGAA 4014
QY 391 IleGluArgCysLysGluMetIleLysAlaGlnGlnGlnLeuGlnGlnLeuAla 410
Db 4015 GCGAACTTAAGAGAGTACGAGTACAGCAGAGGCGTTTCAAAAATCAA----- 4065
QY 411 ThrThrCysAspAspThrThrSerLeuLeuArgAspCysTyrLeuLeuGluGluLys 430
Db 4066 -----GAAGACCATACTTCCAAT-----CCATTAACTGAAGAAAT 4101
QY 431 GluArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysAsnPheGlu----- 448
```

```
Db 4102 CAATTACTTAAGAGGAACTTCAAGTGTTTAAAGAAACCAAGTGACTCTTCTGACATTATT 4161
QY 449 ArgGluArgArgSerPheThrGluAlaAlaIleArgLeuGlyLeuAlaRglyAlaPhe 468
Db 4162 GAGAGAATGAAGAAATTTATGGAGGCGAGAAAGCCCAAGACCATTGAGAAAGGCAACT 4221
QY 469 GluGluGluArg----- 472
Db 4222 GAGTTTGAAGGAACTTAGAAGAAATTTACCGGAAAAAATAGCGAGGAAAGGTAGACAAC 4281
QY 473 -----AlaSerTrpValLysGlnGlnPheLeuAsnMetThr 484
Db 4282 GTGAAATATTGAGGAACTGAAGAAACAATGCTCAAGCAA----- 4323
QY 485 AsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSerGlySerSer 504
Db 4324 ---TATGAAGAGAAACAGTGAAGAAATTA---GAA 4356
QY 505 AspProAspAsnLeuIleValHisSerArg---ProArgGlnLysLysLeuHisSerVal 523
Db 4357 GCGGAAGAGAACTTTAAAGAAAGAAATCAGGTTACCAAGCGCAAGAAAGAAATTCAAAATC 4416
QY 524 AlaAsnGlyValProAlaCysThrSerLysLeuThrLysSerLeuProAlaSerProSer 543
Db 4417 -----ATTAGCAAGCGTAAAGAAAGAAATTTG----- 4440
QY 544 ThrSerAspPheArgGlnThrHisSerCysValSerGluHisSerIleSerValLeu 563
Db 4441 GAACAAGAGTTCCAAAGGAAG-----TTAGAAGAGAAATTCAGGTTGCTTTACATTG 4491
QY 564 AsnIleThrProGluSerLysProSerGluValAlaArgGluSerThrAspGlnLys 583
Db 4492 TCT-----AGCAACAAAAAGGAAACGAAATTAAGTGTGAA-----GACGAAC 4536
QY 584 TrpSerValGlnSerArgProSerSerArgGlu 594
Db 4537 TGAATAGTCCCTCGAAGGGGAAATTCGGAGAAG 4569
```

## RESULT 14

```
US-10-953-349-20044
; Sequence 20044, Application US/10953349
; Publication No. US20060107345A1
; GENERAL INFORMATION:
; APPLICANT: ALEXANDROV, Nikolai et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE OF INVENTION: ENCODED THERBY
; CURRENT APPLICATION NUMBER: US/10/953,349
; CURRENT FILING DATE: 2004-09-30
; NUMBER OF SEQ ID NOS: 40252
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 20044
; LENGTH: 1648
; TYPE: DNA
; ORGANISM: Glycine max
US-10-953-349-20044
```

```
Alignment Scores:
Pred. No.: 4,08e-08 Length: 1648
Score: 188.00 Matches: 118
Percent Similarity: 38.3% Conservative: 100
Best Local Similarity: 20.7% Mismatches: 177
Query Match: 5.9% Indels: 174
DB: 26
```

US-10-644-084-2 (1-615) x US-10-953-349-20044 (1-1648)

```
QY 16 AsnLysAsnLeuSerGlnTyrThrSerGluThrLysMetSerProSerSerLeuTyrSer 35
Db 190 AACGCTGAACACTCAGTCAGGAGTTTCAAAAGATGGAAGGATAAAACCACTCTTTATGAG 249
QY 36 GlnGlnValLeuCysSerSerValProLeuSerLysAsnValHisGlyValPheGlyVal 55
```



Db 897 ---CCTCCAGCTTCTTATGTAAAGACTGCGAGTTCCTTCAAAGAAATACAGAAATTAT--- 950  
QY 54 GlyValPheCysThrGlyGluAsnIleGluGlnSerIleSerTyrLeuAspGlnGluLeu 73  
Db 951 ---GGGAGATTCTCTGAGATGTCAGTGTAT---GAAAAGGAAGTC 992  
QY 74 ThrThrPheGlyPhe-----ProSerLeuTyrGluGluSerIleSer----- 87  
Db 993 ACAGCAGAGGGTGTGAGAGGCCAGAAATTTGCTCAACTTGGTCTTCGGCAGGCATTTCC 1052  
QY 88 -----LysGluAlaLysArgGluLeuAsnIleValAlaValLeuAsnCys---MetAsn 104  
Db 1053 TGGAGGAGTGAACATGTCGGAG-----AACTGTGNGATGCCT 1091  
QY 105 GluLeuValLeuGlnArgLysAsnLeuLeuAlaGlnGluSerValGluThrGlnAsn 124  
Db 1092 GACTGG-----GAGCAAGTCTGAAAGCTTACAA 1121  
QY 125 LeuLysLeuGlySerAspMetAspHisLeuGlnSerCysTyrAlaLysLeuLysGluGln 144  
Db 1122 ---CCTGTTCAAGAGACATG---GCTTTAAATGAAGTCTTGCAGAAATTAATA----- 1169  
QY 145 LeuGluThrSerArgArgGluMetIleGlyLeuGlnGluArgAspArgGlnLeuGlnCys 164  
Db 1170 ---CATACTACAGAAGCAGGAGTGCAGATCCAAAGAA-----CTCCAGTGT 1214  
QY 165 LysAsnArgSerLeuHisGlnLeuLysAsnGluLysAspGluValGlnLysLeuGln 184  
Db 1215 AGTAACCTGATTAGAGAAGGGTTAAAGAACTACAGATGAAGATTACCAACAGCAA 1274  
QY 185 AsnIleAlaSerArgAlaThrGlnTyrAsnHisAspValLysArgLysGluArgGlu 204  
Db 1275 GTGTTCAATT-----GATGTC----- 1289  
QY 205 TyrAsnLysLeuLysGluArgLeuHisGlnLeuValMetAsnLys-----Lys 220  
Db 1290 ATCAATAGCTAAAGAGGAATGTTGAGAAATTAATTGAAGACAAATACAAAATAATCCTA 1349  
QY 221 AspLysAsnIleAlaMetAspValLeuAsnTyrValGlyArgAlaAspGlyLysArgGly 240  
Db 1350 GAGAAGAAAT-----CATACTAAAGACATTTGCAGAAATTTGGAAGAGGTTTAGCT 1400  
QY 241 SerTrpArgThrAspLysThrGluAlaArgAsnGluAspGluMet-----Tyr 256  
Db 1401 AACACGCAAAACATCTTCAGGAATCCAGGAATGCAGGAATAATGTTACAGCTTCAATTT 1460  
QY 257 LysIleLeuLeuAsnAspTyrGluTyrArgGlnLysGlnIleLeuMetGluAsnAlaGlu 276  
Db 1461 AAGAAGATCAAGCTTAATTTATGTGTGTTTACAGGAAGGTACATGACTGAAATGCAACAA 1520  
QY 277 LeuLysLysValLeuGlnGln-----MetLysLysGlu 287  
Db 1521 AAAAATAAATCTGTAAGTCAGTATTTAGAGATGGACAAACCTTAAGCAAGAAGAGAA 1580  
QY 288 MetIleSerLeuLeuSerProGlnLysLysLysProArgGluArgAlaGluAspGlyThr 307  
Db 1581 GAGGTAGAGAGACTACAA---CAACTCAAAAAAGAACTCGAAAAAGGCC----- 1625  
QY 308 GlyThrValAlaIleSerAspIleGluAspAspSerGlyGluLeuSerArgAspSerVal 327  
Db 1626 ---ACAGCTTCTGCTTGGACTTGTGTAACCGGGAAGAAAGAGGCCCAAGAACAGAGTTC 1682  
QY 328 TrpGlyLeuSerCysAspThrValArgGluGlnLeuThrAsnSerIleArgLysGlnTrp 347  
Db 1683 -----TTGCTTTACAGGAGGAATTCAGAAACTTGAAGAGAAACCTGGAAGAAAGA 1736  
QY 348 ArgIleLeuLysSerHisValGluLysLeuAspAsnGlnAlaSerLysValHisSerGlu 367  
Db 1737 CAGAAACTGAATCTAGACTTGAAGAAATTTGCTCACTCAAGTTAGAAATTTGCAATTTATG 1796  
QY 368 GlyLeuAsnGluGluAsp-----ValIleSerArgGlnAspHisGlu-----Gln 382

Db 1797 TCTGAAATGTAAAGAACGAAAGAAATATATAAACTTTACAGCAAAATCAACGAGTAAATAAT 1856  
QY 383 GluThrGluLysLeuGluLeuGluIleGluArgCysLysGlu-----MetIleLys 399  
Db 1857 GAGAAATGCAAAACTTAAACAGCAGAGTTGCAAGGAGTGAAGAGCAAAATATATGTCCTTAA 1916  
QY 400 AlaGlnGlnGlnLeuLeuGlnGlnLeuAlaThrThrCysAspAspAspThrThr--- 418  
Db 1917 TTTGAGACAGCTCAGTTAAAGGATCAATTTAGAGGAAGTCTTGAAGTCAGATATATTACCAAG 1976  
QY 419 -----SerLeuLeuArgAspCysTyrLeuLeuGluGluLysGlu 431  
Db 1977 GATACAAAAACACACATCTCTAATCTGCTCCCGATTGCCACCTTGTGAA-----GAG 2030  
QY 432 ArgLeuLysGluGluTrpThrLeuPheLysGluGlnLysAsnPheGluArgGluArg 451  
Db 2031 AGGCTGAATCCT-----GCAGATATAAAAGAGCTTCTCAGCTGGCTCC 2075  
QY 452 ArgSerPheThrGluAlaAlaIleArgLeuGlyLeu----- 463  
Db 2076 AAAATGCACAGTCTTCTGCTCTGATGTGGAGCTTCTCACATGCCAGGACATCATCAAT 2135  
QY 464 ---GluArgLysAlaPheGluGlu---GluArgAlaSerTrpValLysGlnGlnPheLeu 481  
Db 2136 TCTGATGCTGAACATTTCAAAGAGAGTGAAGAGGTTAGTATATATATGCTGCAAAACTG 2195  
QY 482 AsnMetThrAsnPheAspHisGlnAsnSerGluAsnValLysLeuPheSerAlaPheSer 501  
Db 2196 AAGAGCTCCATCTTAAAAAGAAACTTTAGAT----- 2228  
QY 502 GlySerSerAspProAspAsnLeuIleValHisSerArg----- 514  
Db 2229 -----GAAGAGCTACTGAAACATAAAGATAGAAATCACAACTTTAGAGAG 2273  
QY 515 -----ProArgGlnLysLysLeuHisSerValAlaAsnGlyValProAlaCysThrSer 532  
Db 2274 TTAATTTGTAAGAAAAAGCATTTTCAAGATCATGCTATTAAAGTTCATAGACTGTGATTCA 2333  
QY 533 LysLeuThrLysSerLeuProAlaSerProSer----- 543  
Db 2334 GATGAAGCAAGAGTATCAGAGATGTACCTACCTTCTGGGAGCCAAACTGGATAAGTAC 2393  
QY 544 -----ThrSerAspPheArgGlnThrHisSerCys 553  
Db 2394 CACAGTCTAAATGAGGAGCTTGATTTCTTGGTTACATCATATCAAGAAATCATTTGAATGT 2453  
QY 554 ValSerGluHisSerSerIleSerValLeuAsnIleThrProGluGluSer----- 570  
Db 2454 GCTGACCAAGGCTTGCAATATCCCACTCCAGATTGCACATTTGGAAGAGAGAAATAA 2513  
QY 571 -----LysProSerGluValAlaArgGluSerThrAspGlnLys 583  
Db 2514 CATTTAGAGGATTTAATTTAGAAAGCCAGAGAAAAAGCCAGAAAAACCAAGATCAAAA--- 2570  
QY 584 TrpSerValGlnSerArgProSerSer 592  
Db 2571 ---AGCTTAGAAAAATCATCCGAAGTCC 2594

Search completed: June 14, 2006, 04:09:52  
Job time : 290 secs

THIS PAGE LEFT BLANK